

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 1.90456 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-2

Perfect score: 587

Sequence: 1 QKIVETRIENHNFYK.....TPSASLEVIKDYLSLKRL 119

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.New.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.5	13.5	860	7	US-11-019-711-59
2	76.5	13.0	550	7	US-11-052-554A-200
3	72	12.3	971	7	US-11-052-554A-3
4	71.5	12.2	1889	7	US-11-102-476-46
5	71	12.1	1732	6	US-10-055-877-147
6	68.5	11.7	134	6	US-10-793-626-438
7	68.5	11.7	1237	6	US-11-052-554A-95
8	68	11.6	5024	6	US-10-793-626-2964
9	67.5	11.5	113	6	US-10-793-626-2950
10	67.5	11.5	212	6	US-10-793-626-1622
11	67.5	11.5	341	6	US-10-793-626-226
12	67.5	11.5	532	6	US-10-793-626-546
13	67.5	11.5	583	6	US-10-793-626-1358
14	67	11.4	617	6	US-10-982-545-2
15	67	11.4	968	7	US-11-000-463-281
16	67	11.4	989	6	US-10-821-234-975
17	67	11.4	3635	7	US-11-019-711-47
18	66.5	11.3	857	7	US-11-052-554A-218
19	66.5	11.3	1189	7	US-11-074-176-134
20	66	11.2	477	7	US-11-089-551A-34
21	66	11.2	654	7	US-11-120-308-52
22	65.5	11.2	492	7	US-11-152-697-2
23	65.5	11.2	539	7	US-11-152-697-5
24	65.5	11.2	1424	7	US-11-102-476-4
25	65.5	11.2	1562	7	US-11-052-554A-211

26	65	11.1	1255	7	US-11-022-562-235
27	65	11.1	1255	7	US-11-052-554A-265
28	65	11.1	1255	7	US-11-052-554A-266
29	65	11.1	1255	7	US-11-052-554A-267
30	65	11.1	1255	7	US-11-052-554A-268
31	65	11.1	1255	7	US-11-052-554A-269
32	65	11.1	1255	7	US-11-052-554A-270
33	65	11.1	1255	7	US-11-052-554A-271
34	65	11.1	1255	7	US-11-052-554A-272
35	65	11.1	1255	7	US-11-052-554A-273
36	65	11.1	1255	7	US-11-052-554A-274
37	65	11.1	1255	7	US-11-052-554A-275
38	65	11.1	1279	6	US-10-957-880-3
39	64.5	11.0	222	6	US-10-793-626-128
40	64.5	11.0	593	7	US-11-120-308-54
41	64.5	11.0	3194	7	US-11-052-554A-90
42	64	10.9	495	6	US-10-821-234-1154
43	64	10.9	513	6	US-10-485-517-160
44	64	10.9	1032	7	US-11-014-367-2
45	64	10.9	1452	6	US-10-995-561-778

ALIGNMENTS

RESULT 1

US-11-019-711-59

; Sequence 59, Application US/11019711

; Publication No. US2006009634A1

; GENERAL INFORMATION:

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Alsobrook II, John P

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Liu, Xiaohong

; APPLICANT: Svytek, Kimberly A

; APPLICANT: Patvarajan, Meera

; APPLICANT: Grosse, William M

; APPLICANT: Lepley, Denise M

; APPLICANT: Burgess, Catherine E

; APPLICANT: Vernet, Corine A.M.

; APPLICANT: Li, Li

; APPLICANT: Gorman, Linda

; APPLICANT: Edinger, Shlomit R

; APPLICANT: Sciore, Paul

; APPLICANT: Ellerman, Karen

; APPLICANT: Malyankar, Uriel M

; APPLICANT: Rothenberg, Mark

; APPLICANT: Stone, David J

; APPLICANT: Boldog, Ferenc L

; APPLICANT: Guo, Xiaojia

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Anderson, David W

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Taupier Jr, Raymond J

; APPLICANT: Miller, Charles E

; APPLICANT: Eisen, Andrew J

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-235

; CURRENT APPLICATION NUMBER: US/11/019,711

; CURRENT FILING DATE: 2004-12-21

; PRIOR APPLICATION NUMBER: US/10/037,417

; PRIOR FILING DATE: 2002-09-20

; PRIOR APPLICATION NUMBER: 60/260,018

; PRIOR FILING DATE: 2001-01-05

; PRIOR APPLICATION NUMBER: 60/260,360

; PRIOR FILING DATE: 2001-01-08

; PRIOR APPLICATION NUMBER: 60/272,411

; PRIOR FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 60/272,817

; PRIOR FILING DATE: 2001-03-02

; PRIOR APPLICATION NUMBER: 60/291,186

; PRIOR FILING DATE: 2001-05-15

; PRIOR APPLICATION NUMBER: 60/303,231

Sequence 235, App
Sequence 265, App
Sequence 266, App
Sequence 267, App
Sequence 268, App
Sequence 269, App
Sequence 270, App
Sequence 271, App
Sequence 272, App
Sequence 273, App
Sequence 274, App
Sequence 275, App
Sequence 276, App
Sequence 277, App
Sequence 278, App
Sequence 279, App
Sequence 280, App
Sequence 281, App
Sequence 282, App
Sequence 283, App
Sequence 284, App
Sequence 285, App
Sequence 286, App
Sequence 287, App
Sequence 288, App
Sequence 289, App
Sequence 290, App
Sequence 291, App
Sequence 292, App
Sequence 293, App
Sequence 294, App
Sequence 295, App
Sequence 296, App
Sequence 297, App
Sequence 298, App
Sequence 299, App
Sequence 300, App
Sequence 301, App
Sequence 302, App
Sequence 303, App
Sequence 304, App
Sequence 305, App
Sequence 306, App
Sequence 307, App
Sequence 308, App
Sequence 309, App
Sequence 310, App
Sequence 311, App
Sequence 312, App
Sequence 313, App
Sequence 314, App
Sequence 315, App
Sequence 316, App
Sequence 317, App
Sequence 318, App
Sequence 319, App
Sequence 320, App
Sequence 321, App
Sequence 322, App
Sequence 323, App
Sequence 324, App
Sequence 325, App
Sequence 326, App
Sequence 327, App
Sequence 328, App
Sequence 329, App
Sequence 330, App
Sequence 331, App
Sequence 332, App
Sequence 333, App
Sequence 334, App
Sequence 335, App
Sequence 336, App
Sequence 337, App
Sequence 338, App
Sequence 339, App
Sequence 340, App
Sequence 341, App
Sequence 342, App
Sequence 343, App
Sequence 344, App
Sequence 345, App
Sequence 346, App
Sequence 347, App
Sequence 348, App
Sequence 349, App
Sequence 350, App
Sequence 351, App
Sequence 352, App
Sequence 353, App
Sequence 354, App
Sequence 355, App
Sequence 356, App
Sequence 357, App
Sequence 358, App
Sequence 359, App
Sequence 360, App
Sequence 361, App
Sequence 362, App
Sequence 363, App
Sequence 364, App
Sequence 365, App
Sequence 366, App
Sequence 367, App
Sequence 368, App
Sequence 369, App
Sequence 370, App
Sequence 371, App
Sequence 372, App
Sequence 373, App
Sequence 374, App
Sequence 375, App
Sequence 376, App
Sequence 377, App
Sequence 378, App
Sequence 379, App
Sequence 380, App
Sequence 381, App
Sequence 382, App
Sequence 383, App
Sequence 384, App
Sequence 385, App
Sequence 386, App
Sequence 387, App
Sequence 388, App
Sequence 389, App
Sequence 390, App
Sequence 391, App
Sequence 392, App
Sequence 393, App
Sequence 394, App
Sequence 395, App
Sequence 396, App
Sequence 397, App
Sequence 398, App
Sequence 399, App
Sequence 400, App
Sequence 401, App
Sequence 402, App
Sequence 403, App
Sequence 404, App
Sequence 405, App
Sequence 406, App
Sequence 407, App
Sequence 408, App
Sequence 409, App
Sequence 410, App
Sequence 411, App
Sequence 412, App
Sequence 413, App
Sequence 414, App
Sequence 415, App
Sequence 416, App
Sequence 417, App
Sequence 418, App
Sequence 419, App
Sequence 420, App
Sequence 421, App
Sequence 422, App
Sequence 423, App
Sequence 424, App
Sequence 425, App
Sequence 426, App
Sequence 427, App
Sequence 428, App
Sequence 429, App
Sequence 430, App
Sequence 431, App
Sequence 432, App
Sequence 433, App
Sequence 434, App
Sequence 435, App
Sequence 436, App
Sequence 437, App
Sequence 438, App
Sequence 439, App
Sequence 440, App
Sequence 441, App
Sequence 442, App
Sequence 443, App
Sequence 444, App
Sequence 445, App
Sequence 446, App
Sequence 447, App
Sequence 448, App
Sequence 449, App
Sequence 450, App
Sequence 451, App
Sequence 452, App
Sequence 453, App
Sequence 454, App
Sequence 455, App
Sequence 456, App
Sequence 457, App
Sequence 458, App
Sequence 459, App
Sequence 460, App
Sequence 461, App
Sequence 462, App
Sequence 463, App
Sequence 464, App
Sequence 465, App
Sequence 466, App
Sequence 467, App
Sequence 468, App
Sequence 469, App
Sequence 470, App
Sequence 471, App
Sequence 472, App
Sequence 473, App
Sequence 474, App
Sequence 475, App
Sequence 476, App
Sequence 477, App
Sequence 478, App
Sequence 479, App
Sequence 480, App
Sequence 481, App
Sequence 482, App
Sequence 483, App
Sequence 484, App
Sequence 485, App
Sequence 486, App
Sequence 487, App
Sequence 488, App
Sequence 489, App
Sequence 490, App
Sequence 491, App
Sequence 492, App
Sequence 493, App
Sequence 494, App
Sequence 495, App
Sequence 496, App
Sequence 497, App
Sequence 498, App
Sequence 499, App
Sequence 500, App
Sequence 501, App
Sequence 502, App
Sequence 503, App
Sequence 504, App
Sequence 505, App
Sequence 506, App
Sequence 507, App
Sequence 508, App
Sequence 509, App
Sequence 510, App
Sequence 511, App
Sequence 512, App
Sequence 513, App
Sequence 514, App
Sequence 515, App
Sequence 516, App
Sequence 517, App
Sequence 518, App
Sequence 519, App
Sequence 520, App
Sequence 521, App
Sequence 522, App
Sequence 523, App
Sequence 524, App
Sequence 525, App
Sequence 526, App
Sequence 527, App
Sequence 528, App
Sequence 529, App
Sequence 530, App
Sequence 531, App
Sequence 532, App
Sequence 533, App
Sequence 534, App
Sequence 535, App
Sequence 536, App
Sequence 537, App
Sequence 538, App
Sequence 539, App
Sequence 540, App
Sequence 541, App
Sequence 542, App
Sequence 543, App
Sequence 544, App
Sequence 545, App
Sequence 546, App
Sequence 547, App
Sequence 548, App
Sequence 549, App
Sequence 550, App
Sequence 551, App
Sequence 552, App
Sequence 553, App
Sequence 554, App
Sequence 555, App
Sequence 556, App
Sequence 557, App
Sequence 558, App
Sequence 559, App
Sequence 560, App
Sequence 561, App
Sequence 562, App
Sequence 563, App
Sequence 564, App
Sequence 565, App
Sequence 566, App
Sequence 567, App
Sequence 568, App
Sequence 569, App
Sequence 570, App
Sequence 571, App
Sequence 572, App
Sequence 573, App
Sequence 574, App
Sequence 575, App
Sequence 576, App
Sequence 577, App
Sequence 578, App
Sequence 579, App
Sequence 580, App
Sequence 581, App
Sequence 582, App
Sequence 583, App
Sequence 584, App
Sequence 585, App
Sequence 586, App
Sequence 587, App
Sequence 588, App
Sequence 589, App
Sequence 590, App
Sequence 591, App
Sequence 592, App
Sequence 593, App
Sequence 594, App
Sequence 595, App
Sequence 596, App
Sequence 597, App
Sequence 598, App
Sequence 599, App
Sequence 600, App
Sequence 601, App
Sequence 602, App
Sequence 603, App
Sequence 604, App
Sequence 605, App
Sequence 606, App
Sequence 607, App
Sequence 608, App
Sequence 609, App
Sequence 610, App
Sequence 611, App
Sequence 612, App
Sequence 613, App
Sequence 614, App
Sequence 615, App
Sequence 616, App
Sequence 617, App
Sequence 618, App
Sequence 619, App
Sequence 620, App
Sequence 621, App
Sequence 622, App
Sequence 623, App
Sequence 624, App
Sequence 625, App
Sequence 626, App
Sequence 627, App
Sequence 628, App
Sequence 629, App
Sequence 630, App
Sequence 631, App
Sequence 632, App
Sequence 633, App
Sequence 634, App
Sequence 635, App
Sequence 636, App
Sequence 637, App
Sequence 638, App
Sequence 639, App
Sequence 640, App
Sequence 641, App
Sequence 642, App
Sequence 643, App
Sequence 644, App
Sequence 645, App
Sequence 646, App
Sequence 647, App
Sequence 648, App
Sequence 649, App
Sequence 650, App
Sequence 651, App
Sequence 652, App
Sequence 653, App
Sequence 654, App
Sequence 655, App
Sequence 656, App
Sequence 657, App
Sequence 658, App
Sequence 659, App
Sequence 660, App
Sequence 661, App
Sequence 662, App
Sequence 663, App
Sequence 664, App
Sequence 665, App
Sequence 666, App
Sequence 667, App
Sequence 668, App
Sequence 669, App
Sequence 670, App
Sequence 671, App
Sequence 672, App
Sequence 673, App
Sequence 674, App
Sequence 675, App
Sequence 676, App
Sequence 677, App
Sequence 678, App
Sequence 679, App
Sequence 680, App
Sequence 681, App
Sequence 682, App
Sequence 683, App
Sequence 684, App
Sequence 685, App
Sequence 686, App
Sequence 687, App
Sequence 688, App
Sequence 689, App
Sequence 690, App
Sequence 691, App
Sequence 692, App
Sequence 693, App
Sequence 694, App
Sequence 695, App
Sequence 696, App
Sequence 697, App
Sequence 698, App
Sequence 699, App
Sequence 700, App
Sequence 701, App
Sequence 702, App
Sequence 703, App
Sequence 704, App
Sequence 705, App
Sequence 706, App
Sequence 707, App
Sequence 708, App
Sequence 709, App
Sequence 710, App
Sequence 711, App
Sequence 712, App
Sequence 713, App
Sequence 714, App
Sequence 715, App
Sequence 716, App
Sequence 717, App
Sequence 718, App
Sequence 719, App
Sequence 720, App
Sequence 721, App
Sequence 722, App
Sequence 723, App
Sequence 724, App
Sequence 725, App
Sequence 726, App
Sequence 727, App
Sequence 728, App
Sequence 729, App
Sequence 730, App
Sequence 731, App
Sequence 732, App
Sequence 733, App
Sequence 734, App
Sequence 735, App
Sequence 736, App
Sequence 737, App
Sequence 738, App
Sequence 739, App
Sequence 740, App
Sequence 741, App
Sequence 742, App
Sequence 743, App
Sequence 744, App
Sequence 745, App
Sequence 746, App
Sequence 747, App
Sequence 748, App
Sequence 749, App
Sequence 750, App
Sequence 751, App
Sequence 752, App
Sequence 753, App
Sequence 754, App
Sequence 755, App
Sequence 756, App
Sequence 757, App
Sequence 758, App
Sequence 759, App
Sequence 760, App
Sequence 761, App
Sequence 762, App
Sequence 763, App
Sequence 764, App
Sequence 765, App
Sequence 766, App
Sequence 767, App
Sequence 768, App
Sequence 769, App
Sequence 770, App
Sequence 771, App
Sequence 772, App
Sequence 773, App
Sequence 774, App
Sequence 775, App
Sequence 776, App
Sequence 777, App
Sequence 778, App
Sequence 779, App
Sequence 780, App
Sequence 781, App
Sequence 782, App
Sequence 783, App
Sequence 784, App
Sequence 785, App
Sequence 786, App
Sequence 787, App
Sequence 788, App
Sequence 789, App
Sequence 790, App
Sequence 791, App
Sequence 792, App
Sequence 793, App
Sequence 794, App
Sequence 795, App
Sequence 796, App
Sequence 797, App
Sequence 798, App
Sequence 799, App
Sequence 800, App
Sequence 801, App
Sequence 802, App
Sequence 803, App
Sequence 804, App
Sequence 805, App
Sequence 806, App
Sequence 807, App
Sequence 808, App
Sequence 809, App
Sequence 810, App
Sequence 811, App
Sequence 812, App
Sequence 813, App
Sequence 814, App
Sequence 815, App
Sequence 816, App
Sequence 817, App
Sequence 818, App
Sequence 819, App
Sequence 820, App
Sequence 821, App
Sequence 822, App
Sequence 823, App
Sequence 824, App
Sequence 825, App
Sequence 826, App
Sequence 827, App
Sequence 828, App
Sequence 829, App
Sequence 830, App
Sequence 831, App
Sequence 832, App
Sequence 833, App
Sequence 834, App
Sequence 835, App
Sequence 836, App
Sequence 837, App
Sequence 838, App
Sequence 839, App
Sequence 840, App
Sequence 841, App
Sequence 842, App
Sequence 843, App
Sequence 844, App
Sequence 845, App
Sequence 846, App
Sequence 847, App
Sequence 848, App
Sequence 849, App
Sequence 850, App
Sequence 851, App
Sequence 852, App
Sequence 853, App
Sequence 854, App
Sequence 855, App
Sequence 856, App
Sequence 857, App
Sequence 858, App
Sequence 859, App
Sequence 860, App
Sequence 861, App
Sequence 862, App
Sequence 863, App
Sequence 864, App
Sequence 865, App
Sequence 866, App
Sequence 867, App
Sequence 868, App
Sequence 869, App
Sequence 870, App
Sequence 871, App
Sequence 872, App
Sequence 873, App
Sequence 874, App
Sequence 875, App
Sequence 876, App
Sequence 877, App
Sequence 878, App
Sequence 879, App
Sequence 880, App
Sequence 881, App
Sequence 882, App
Sequence 883, App
Sequence 884, App
Sequence 885, App

[illegible]

Sequence 147, Application US/10055877
Publication No. US20050288241A1
GENERAL INFORMATION:
APPLICANT: DeCristofaro, Marc
APPLICANT: Padigaru, Muralidhara
APPLICANT: Miller, Charles
APPLICANT: Tchernev, Velizar
APPLICANT: Zhong, Mei
APPLICANT: Anderson, David
APPLICANT: Ballinger, Robert
APPLICANT: Gerlach, Valerie
APPLICANT: Spytek, Kimberly
APPLICANT: Ratelli, Luca
APPLICANT: Kekuda, Ramesh
APPLICANT: Guo, Xiaojia
APPLICANT: Zerhusen, Bryan
APPLICANT: Andrew, David
APPLICANT: Mezes, Peter
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine
APPLICANT: Eisen, Andrew
APPLICANT: Wolenc, Adam
APPLICANT: Baumgartner, Jason
APPLICANT: Shimkets, Richard
APPLICANT: Gusev, Vladimir
APPLICANT: Vernet, Corine
APPLICANT: Taupier Jr., Raymond
APPLICANT: Pena, Carol
APPLICANT: Shenoy, Suresh
APPLICANT: Li, Li
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ference
TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby
FILE REFERENCE: 21402-251
CURRENT APPLICATION NUMBER: US/10/055,877
CURRENT FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: 60/262,892
PRIOR FILING DATE: 2001-01-19
PRIOR APPLICATION NUMBER: 60/263,598
PRIOR FILING DATE: 2001-01-23
PRIOR APPLICATION NUMBER: 60/263,799
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/264,117
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/264,139
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/264,478
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/263,351
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/272,870
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: 60/275,990
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/275,927
PRIOR FILING DATE: 2001-03-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 512
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 147
LENGTH: 1732
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-055-877-147

Query Match 12.1%; Score 71; DB 6; Length 1732;
Best Local Similarity 24.1%; Pred. No. 49;
Matches 33; Conservative 25; Mismatches 53; Indels 26; Gaps 5;
Qy 1 OKIYVEIERRIENHFLFKDSLVQLQDALSSATTSI-----SALTQSNDRGSGLLSS 55
Db 594 OKLARHVRDKKEEVLVMQAESEQLRRAERAKKELEVHTTEALTAASAKDRK----- 647

Qy 56 FLRKQNNHNSKDISN-----LRTLNDLSL-----QELARLKSNNNEGMFYATPS 102
Db 648 -LREQ-SRHSYKQDENELEGLKQKQISYSPGICSIHQOEITKLTDLKKSIFYEEIS 705
Qy 103 ASLEVIKYDLSYLRRL 119
Db 706 KREGIHASEIKNLKEL 722
RESULT 6
US-10-793-626-438
Sequence 438, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 438
LENGTH: 134
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-10-793-626-438

Query Match 11.7%; Score 68.5; DB 6; Length 134;
Best Local Similarity 20.9%; Pred. No. 3.3;
Matches 19; Conservative 23; Mismatches 40; Indels 9; Gaps 2;
Qy 18 FYKDESLVQLQDALSSATTSLSALTQSNDRGSGLLSFLRKQNNHNSKDISNLT--- 74
Db 35 FGRQEQQLILEANMADKMTSASSITHFNT-----LMNQIKNVKTDLGLDLTIRTKYK 89
Qy 75 -LNDLSLQELARLKSNNNEGMFYATPSAS 104
Db 90 DANDQVNRHQLEGGEGGQNDGLYFIIPSDAS 120

RESULT 7
US-11-052-554A-95
Sequence 95, Application US/11052554A
Publication No. US20050288666A1
GENERAL INFORMATION:
APPLICANT: Sachdeva, et al.
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
FILE REFERENCE: 30853/40359A
CURRENT APPLICATION NUMBER: US/11/052,554A
CURRENT FILING DATE: 2005-02-07
PRIOR APPLICATION NUMBER: US 60/589,227
PRIOR FILING DATE: 2004-07-20
PRIOR APPLICATION NUMBER: IN 173/DEL/2004
PRIOR FILING DATE: 2004-02-06
NUMBER OF SEQ ID NOS: 763
SOFTWARE: PatentIn version 3.3
SEQ ID NO 95
LENGTH: 1237
TYPE: PRT
ORGANISM: Helicobacter pylori J99
US-11-052-554A-95

Query Match 11.7%; Score 68.5; DB 7; Length 1237;
Best Local Similarity 24.6%; Pred. No. 56;
Matches 33; Conservative 17; Mismatches 47; Indels 37; Gaps 4;
Qy 6 BIERRIENHFLFYKDESLVQLQDAL---SSATTSLSALTQSNDRGSGLL----- 53


```

RESULT 15
US-11-000-463-281
/ Sequence 281, Application US/11000463
/ Publication No. US20050266423A1
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Qian, Xiaohong B.
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Zhang, Jie
/ APPLICANT: Zhou, Ping
/ APPLICANT: Cao, Yi-Cheng
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 785C1P4CN
/ CURRENT APPLICATION NUMBER: US/11/000,463
/ CURRENT FILING DATE: 2004-11-29
/ PRIOR APPLICATION NUMBER: 10/291,265
/ PRIOR FILING DATE: 2002-11-08
/ PRIOR APPLICATION NUMBER: PCT/US01/02623
/ PRIOR FILING DATE: 2001-01-25
/ PRIOR APPLICATION NUMBER: 09/922,279
/ PRIOR FILING DATE: 2001-08-03
/ PRIOR APPLICATION NUMBER: 09/491,404
/ PRIOR FILING DATE: 2000-01-25
/ PRIOR APPLICATION NUMBER: 09/617,746
/ PRIOR FILING DATE: 2000-07-17
/ PRIOR APPLICATION NUMBER: 09/631,451
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: 09/633,870
/ PRIOR FILING DATE: 2000-09-15
/ NUMBER OF SEQ ID NOS: 944
/ SOFTWARE: Fast-Seq for Windows Version 3.0
/ SEQ ID NO 281
/ LENGTH: 968
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-11-000-463-281

```

Query Match	11.4%;	Score 67;	DB 7;	Length 968;
Best Local Similarity	23.3%;	Pred. No. 57;		
Matches	41;	Conservative 25;	Mismatches 50;	Indels 60; Gaps 8;

Qy	1	QKIYVEIRERIEHNHFLFYKDESLV---	QLQDALSSATTSLTSALTQSNDDR--	GSGLSSSF	56
		: : : : :	: : : :	: : : :	
Db	445	EDVALEETIYENEGFRPYAEDTUSARGSGSDSPGSSSLU--	ITRKNSPKSGSPKSSSL	503	
		: : : : :	: : : :	: : : :	
Qy	57	LR-----	KQNSNNHSDKISNLRITLN-----	DSLQELARLKSNLN-----	91
		: : : : :	: : : :	: : : :	
Db	504	LKLKAEKNAQAEWGKHNSSASFSSSIITNTCCSSSSSSSSSLSKTSGDLKPRASDAGI	563		
		: : : : :	: : : :	: : : :	
Qy	92	-----	NEGMYTATPSASLEV-----	IKYDSLVLKRL	119
		: : : : :	: : : :	: : : :	
Db	564	RGTPKVRAKKDADANAGL--	TSCPRAPKSVRPKPFNLRAESQSQEKMDISTLRQL	617	
		: : : : :	: : : :	: : : :	

Search completed: January 24, 2006, 20:55:32
Job time : 2.90456 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 20.9149 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-2
Perfect score: 587
Sequence: 1 QKIYVEIERIENHNFYK.....TPSASLEVIKYDLSYLRRL 119

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	587	100.0	119	5	US-10-688-058-2
2	566	96.4	407	5	US-10-688-058-4
3	91	15.5	326	3	US-09-943-002-8
4	90.5	15.4	2910	5	US-10-732-923-3342
5	88	15.0	503	5	US-10-450-763-50636
6	88	15.0	726	5	US-10-483-680-2
7	88	15.0	2400	5	US-10-450-763-50637
8	86	14.7	733	4	US-10-424-599-167699
9	86	14.7	1242	4	US-10-437-963-126385
10	86	14.7	1310	4	US-10-437-963-126423
11	86	14.7	1511	4	US-10-437-963-126422
12	85.5	14.6	523	5	US-10-450-763-50635
13	84	14.3	209	4	US-10-424-599-279739
14	84	14.3	602	5	US-10-732-923-13440
15	84	14.3	1309	5	US-10-450-763-50637
16	84	14.3	1984	6	US-11-097-143-9972
17	83	14.1	650	4	US-10-425-114-49905
18	83	14.1	1091	4	US-10-369-493-6328
19	82.5	14.1	761	4	US-10-416-330-35
20	82	14.0	2415	5	US-10-450-763-50638
21	81.5	13.9	343	4	US-10-425-115-204637
22	81	13.8	1832	4	US-10-437-963-126381
23	81	13.8	4624	4	US-10-408-765A-2991
24	80.5	13.7	666	3	US-09-801-368-36
25	80	13.6	186	4	US-10-724-972A-6870
26	80	13.6	497	4	US-10-425-115-279869
27	79.5	13.5	336	3	US-09-976-782-33

28	79.5	13.5	336	4	US-10-231-913-262	Sequence 262, App
29	79.5	13.5	860	4	US-10-080-334-166	Sequence 166, App
30	79.5	13.5	860	4	US-10-072-012-838	Sequence 838, App
31	79.5	13.5	860	4	US-10-037-417-59	Sequence 59, Appl
32	79.5	13.5	860	5	US-10-494-343-866	Sequence 866, App
33	79	13.5	570	4	US-10-425-114-57234	Sequence 57234, A
34	79	13.5	4080	4	US-10-307-817-138	Sequence 138, App
35	78.5	13.4	350	4	US-10-425-114-72202	Sequence 72202, A
36	78.5	13.4	950	4	US-10-282-122A-47285	Sequence 47285, A
37	78.5	13.4	1043	6	US-11-097-143-41328	Sequence 41328, A
38	78.5	13.4	1669	4	US-10-425-115-291649	Sequence 291649, A
39	78.5	13.4	1819	4	US-10-161-051-97	Sequence 97, Appl
40	78	13.3	589	4	US-10-425-114-68204	Sequence 68204, A
41	78	13.3	754	4	US-10-425-115-298211	Sequence 298211, A
42	78	13.3	4498	4	US-10-712-124-68	Sequence 68, Appl
43	78	13.3	4498	6	US-11-097-143-2577	Sequence 2577, Ap
44	77.5	13.2	918	6	US-11-097-143-6468	Sequence 6468, Ap
45	77.5	13.2	1246	6	US-11-097-143-19635	Sequence 19635, A

ALIGNMENTS

RESULT 1
US-10-688-058-2
; Sequence 2, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:00305
; CURRENT APPLICATION NUMBER: US/10/688.058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-2

Query Match	100.0%	Score	587	DB	5	Length	119
Best Local Similarity	100.0%	Pred. No.	3.7e-53				
Matches	119	Conservative	0	Mismatches	0	Indels	0
QY	1	QKIYVEIERIENHNFYKDES	VLQDALSSATTSLSALTQSNDRGSLSSFLRKQ	60			
DB	1	QKIYVEIERIENHNFYKDES	VLQDALSSATTSLSALTQSNDRGSLSSFLRKQ	60			
QY	61	NSNHSKDISNRLTNDLSQELARKLNKLNNEGFFYATPSASLEVIKYDLSYLRRL	119				
DB	61	NSNHSKDISNRLTNDLSQELARKLNKLNNEGFFYATPSASLEVIKYDLSYLRRL	119				

RESULT 2
US-10-688-058-4
; Sequence 4, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:00305
; CURRENT APPLICATION NUMBER: US/10/688.058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-4

Query Match          96.4%; Score 566; DB 5; Length 407;
Best Local Similarity 95.8%; Pred. No. 3e-50;
Matches 114; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QKIYVEIERRIENHNFYKDESLSVQLQDALSSATTSLTQSNNDRGSGLLSFLRKQ 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 EKIYVEIERRIENHNFYKDESLSVQLQDALSSATTSLTQSNNDRGSGLLSFLRKQ 233
QY 61 NSNNHSDISNRLTNDLSQELARLKSINLNNEGFMFYATPSPASLEVIKYDLSYKRL 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 NSNNHSDISNRLTNDLSQELARLKSINLNNEGFMFYATPSPASLEVIKYDLSYKRL 292

RESULT 3
US-09-943-002-8
; Sequence 8, Application US/09943002
; Patent No. US20020045734A1
; GENERAL INFORMATION:
; APPLICANT: Duncan, Roy
; TITLE OF INVENTION: NOVEL REOVIRUS-DERIVED PROTEINS AND USES THEREFOR
; FILE REFERENCE: 78973-1C
; CURRENT APPLICATION NUMBER: US/09/943,002
; CURRENT FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 326
; TYPE: PRT
; ORGANISM: avian reovirus strain 138
US-09-943-002-8

Query Match          15.5%; Score 91; DB 3; Length 326;
Best Local Similarity 29.5%; Pred. No. 0.64;
Matches 28; Conservative 23; Mismatches 24; Indels 20; Gaps 3;

QY 23 SLVQLQDALSSATTSLTQSNNDRGSGLLSFLRKQNSNNHSDISNRLTNDLSQEL 82
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 TITTLQDSVSTLSTVTDLTNT-----SSVHSEALSRLTIVDGNSTT 127

QY 83 LARLKSINLNNEGFMFYATPSPASLEVIKYDLSY 114
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 IDNLKSDVSSNGLAITDLQSRVKLSLSVSSHGLSF 162

RESULT 4
US-10-732-923-3342
; Sequence 3342, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3342
; LENGTH: 2910
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-3342

Query Match          15.4%; Score 90.5; DB 5; Length 2910;
Best Local Similarity 26.3%; Pred. No. 13;
Matches 30; Conservative 24; Mismatches 47; Indels 13; Gaps 4;
```

```
QY 8 ERRIENHNFYKDESLSVQLQDALSSATTSLTQSNND-----RGSGLLSFLRKQNSN 63
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 EKSISNYNF-----DKTLFLFYNVLLKVSVHISGLYNTFMDILNTYSSDALNNHINSNNN 479
QY 64 NHSKDISNRLTND-SLSQELARLKSINLNNEGFMFYATPSPASLEVIKYDLSYK 116
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
480 NNNNNNNNNNDVFFISYELRKLESIN-----IYQRAVNAALYLFHYLHMLE 529

RESULT 5
US-10-450-763-50636
; Sequence 50636, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 50636
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (267)..(315)
; OTHER INFORMATION: Receptor tyrosine kinase class II proteins domain identified
; OTHER INFORMATION: by eMATRIX, accession number BL00239B, p-value=1.000e-40, raw sc
; OTHER INFORMATION: of 25.15
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (217)..(490)
; OTHER INFORMATION: Eukaryotic protein kinase domain identified by PFam,
; OTHER INFORMATION: accession name pkinase, E-value=9.6e-96, PFam score of 331.5
US-10-450-763-50636

Query Match          15.0%; Score 88; DB 5; Length 503;
Best Local Similarity 29.3%; Pred. No. 2.4;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;

QY 4 YVEIERRIENHNFYKDESLSV-----LQDQDALSSATTSLTQSN-----NDRGS 50
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 YFEIEKR-----LSHQERLVNRETQCQSURLEKLNQKALTEKKNKELEIAQDRNI 107
QY 51 GLLSFLR-KQNSNNHSDISNRLTNDLSQELARLKSINLN--NEGFMFYATPSPASLEV 107
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
108 AIQSOFTTKKEEAKEKDL--IRT-NERLSQELELYTDEVKRLNEKLESNTTKGELQ- 163
QY 108 IKYD-----LSYLKRL 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 LKLDLQASDVSVKYREKRL 183

RESULT 6
US-10-483-680-2
; Sequence 2, Application US/10483680
; Publication No. US20050074839A1
; GENERAL INFORMATION:
; APPLICANT: Merck Patent GmbH
; APPLICANT: Merck Patent GmbH
; TITLE OF INVENTION: Survivin Interacting Protein TPRI
; FILE REFERENCE: TPRI8HWS
; CURRENT APPLICATION NUMBER: US/10/483,680
; CURRENT FILING DATE: 2004-01-13
```

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-483-680-2

Query Match 15.0%; Score 88; DB 5; Length 726;
Best Local Similarity 29.3%; Pred. No. 3.8;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;
Qy 4 YVEIERRIENHFLFYKDESLV-----QLQDALSSATTSLSALTQSN-----NDRGS 50
Db 54 YFEIEKR-----LSHSQERLVNTRQCOSLRLEKLNQKALKTERKKELEIAQDRNI 107
Qy 51 GLLSFLR-KQNSNNHSDINSLRTINDLSQELARLKSINL--NEGMPFYATPSASLEV 107
Db 108 AIQSQFTTRKELEAEKRDL--IRT-NERLSQELELYLTVKRLNEKLSKESNTTKGELQ- 163
Qy 108 IKYD-----LSYLKRRL 119
Db 164 LKDELQASDVSVKYRKRL 183

RESULT 7
US-10-450-763-50637
; Sequence 50637, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 50637
; LENGTH: 2400
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (351)..(362)
; OTHER INFORMATION: 8 kw CHO2 ALPHA ANTIGEN PARAMYOSIN domain identified by
; OTHER INFORMATION: eMATRIX, accession number DM00588A, p-value=5.950e-09, raw score
; OTHER INFORMATION: 10.87
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2400)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-50637

Query Match 15.0%; Score 88; DB 5; Length 2400;
Best Local Similarity 29.3%; Pred. No. 19;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;
Qy 4 YVEIERRIENHFLFYKDESLV-----QLQDALSSATTSLSALTQSN-----NDRGS 50
Db 83 YFEIEKR-----LSHSQERLVNTRQCOSLRLEKLNQKALKTERKKELEIAQDRNI 136
Qy 51 GLLSFLR-KQNSNNHSDINSLRTINDLSQELARLKSINL--NEGMPFYATPSASLEV 107
Db 137 AIQSQFTTRKELEAEKRDL--IRT-NERLSQELELYLTVKRLNEKLSKESNTTKGELQ- 192
Qy 108 IKYD-----LSYLKRRL 119

Db 193 LKDELQASDVSVKYRKRL 212
RESULT 8
US-10-424-599-167699
; Sequence 167699, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 167699
; LENGTH: 733
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(733)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_122449C.1.pap
US-10-424-599-167699

Query Match 14.7%; Score 86; DB 4; Length 733;
Best Local Similarity 23.4%; Pred. No. 6.3;
Matches 40; Conservative 28; Mismatches 49; Indels 54; Gaps 6;
Qy 3 IYVEIERRI-----ENHNF-----LFYKDESLVQLQD 29
Db 323 LVIELNSRVSSLTLENNFRSKCDVMEKEYNVLFSAEKKALDSKVLREKDEELHQLKD 382
Qy 30 ----ALSATTS-----LSALTOSNDRSGLLS-----SFLRKQ-----NSNNHSD 68
Db 383 QFELALGEASKSQIVADIADLSQERDLDKEALDNSSKVNHLKQELQVTOENLAKSRNESAE 442
Qy 69 ISNRLTINDLSQELARLKSINLNEGMPFYATPSASLEVIKYDLSYLKRRL 119
Db 443 LENLLTSLNKLCKELELEVSKLSSELTEVNESLQRLNDLDAKHEAFMLASEL 493

RESULT 9
US-10-437-963-126385
; Sequence 126385, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 126385
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_28937C.1.pap
US-10-437-963-126385

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 7.3361 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-2

Perfect score: 587

Sequence: 1 QKIVETRIENHNFYK.....TPSASLEVIKYLKRL 119

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/PCUTS COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	15.0	2349	2	US-09-538-092-914
2	82	14.0	294	2	US-09-248-796A-25451
3	80.5	13.7	666	2	US-09-487-558B-36
4	80	13.6	186	2	US-09-134-001C-4584
5	79.5	13.5	860	2	US-10-037-417-59
6	79	13.5	1290	2	US-09-248-796A-20654
7	78.5	13.4	333	2	US-09-248-796A-19426
8	78.5	13.4	1627	2	US-09-540-236-3533
9	76.5	13.0	214	2	US-09-248-796A-20889
10	76.5	13.0	367	2	US-09-543-681A-8041
11	76.5	13.0	2807	2	US-09-543-681A-4980
12	75.5	12.9	500	1	US-08-260-582-77
13	75.5	12.9	500	4	PCT-US95-05471-77
14	75.5	12.9	569	2	US-09-248-796A-16697
15	75	12.8	775	1	US-07-603-133B-13
16	75	12.8	1427	2	US-09-538-092-1044
17	74.5	12.7	2037	2	US-09-248-796A-22095
18	74.5	12.7	237	2	US-09-248-796A-18794
19	74.5	12.7	480	2	US-09-248-796A-20303
20	74	12.6	515	2	US-09-270-767-45839
21	74	12.6	733	2	US-09-232-338-10
22	73.5	12.5	353	2	US-09-248-796A-15118
23	73.5	12.5	415	2	US-09-543-681A-8002
24	73.5	12.5	592	2	US-09-861-451A-72
25	73.5	12.5	845	2	US-09-248-796A-17962
26	73	12.4	440	2	US-09-762-194-2
27	73	12.4	440	2	US-09-762-194-4

```

28 72.5 12.4 339 2 US-09-583-110-3067 Sequence 3067, Ap
29 72.5 12.4 348 2 US-09-107-433-2966 Sequence 2966, Ap
30 72.5 12.4 435 2 US-09-762-194-8 Sequence 8, Appli
31 72.5 12.4 435 2 US-09-762-194-10 Sequence 10, Appl
32 72.5 12.4 563 2 US-09-134-001C-3172 Sequence 3172, Ap
33 72.5 12.4 668 1 US-08-468-036-5 Sequence 5, Appli
34 72.5 12.4 668 1 US-08-376-843-5 Sequence 5, Appli
35 72.5 12.4 823 2 US-09-538-092-19 Sequence 19, Appl
36 72.5 12.4 823 2 US-09-248-796A-16699 Sequence 16699, A
37 72.5 12.4 1041 2 US-09-999-833A-498 Sequence 498, App
38 72.5 12.4 1041 2 US-09-954-987B-184 Sequence 184, App
39 72.5 12.4 1041 2 US-09-954-987B-186 Sequence 186, App
40 72.5 12.4 1041 2 US-10-020-445A-498 Sequence 498, App
41 72.5 12.4 1059 2 US-09-954-987B-187 Sequence 187, App
42 72.5 12.4 1111 2 US-09-914-259-28 Sequence 28, Appl
43 72.5 12.4 1177 2 US-09-134-001C-5106 Sequence 5106, Ap
44 72 12.3 290 2 US-09-134-000C-4994 Sequence 4994, Ap
45 72 12.3 580 2 US-09-248-796A-14473 Sequence 14473, A

```

ALIGNMENTS

RESULT 1

```

US-09-538-092-914
; Sequence 914, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Ioic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

```

```

Query Match 15.0%; Score 88; DB 2; Length 2349;
Best Local Similarity 29.3%; Pred. No. 3.1;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;

Qy 4 YVETRIENHNFYKDESLV-----QLQDALSSATTSLTQSN-----NDRGS 50
Db 54 YFELEK-----LSHQERLVNTRCQSLRLEKLNOLKALTEKKNLETAQDRNI 107

Qy 51 GLLSFFLR-KONSNNHSDKISNLTNLSLSQELARKLSNLTN--NEGMYTATPSASLEV 107
Db 108 AIQSQFTRTKEELEAEKRD--INT-NERLSQELELTEDVKRLNEKLKESNTTKGEIQ- 163

Qy 108 IKYD-----LSYKRL 119
Db 164 LKLELQASDVSVKYREKRL 183

```

RESULT 2

```

US-09-248-796A-25451
; Sequence 25451, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al

```

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 25451
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-25451

Query Match 14.0%; Score 82; DB 2; Length 294;
Best Local Similarity 28.4%; Pred. No. 0.77;
Matches 33; Conservative 23; Mismatches 36; Indels 24; Gaps 6;

QY 1 QKIYVEIERRHHNFYKDESIVQLQDALSSATTSLSALTQSNDR-----G 49
DB 142 QTVQSIDSFIDESSVLYVDLEN---EENLSSNSLSIV--NODHTYLQNKIHRGIN 195
QY 50 SGLSSFLRKQNSNNHNSKDISNRLTLN---DSLQELARLK--SNLANE-GMFT 98
DB 196 SYNLSDFVNRNNSNNNDNNANVEKNKSFELDQLQAWRLTGINNVRNVGEFT 251

RESULT 3
US-09-487-558B-36
; Sequence 36, Application US/09487558B
; Patent No. 6949356
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. 6949356man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofia
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.130
; CURRENT APPLICATION NUMBER: US/09/487,558B
; CURRENT FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/487,558
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 446
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 36
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-487-558B-36

Query Match 13.7%; Score 80.5; DB 2; Length 666;
Best Local Similarity 27.0%; Pred. No. 3.5;
Matches 24; Conservative 19; Mismatches 37; Indels 9; Gaps 3;

QY 13 NHHNFYKDES----LVQLQDALSSATTSLSA-----LTOS--NDRGSGLLSFLRKQNSN 63
DB 281 NKNYYKNDSSGYPITQAPALPSTLMYTANPPYITQSPDNTWATGNTHVNNNNNS 340
QY 64 NNSKDISNRLTNDLSQELARLKNLNN 92
DB 341 NNSNSNNNSNNNNNNNNNNNNNNNNNN 369
```

RESULT 4

```
US-09-134-001C-4584
; Sequence 4584, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4584
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4584

Query Match 13.6%; Score 80; DB 2; Length 186;
Best Local Similarity 27.0%; Pred. No. 0.67;
Matches 30; Conservative 27; Mismatches 32; Indels 22; Gaps 6;

QY 1 QKIYVEIERRHHNFYKDESIVQLQD--ALSATTLSALTQSNDRGSGLLSFL- 57
DB 19 KKIYTEIQRLLETHF---RDTIDQISENTGISKAT-----IYRWKDKSSIIMSAFIE 70
QY 58 RKQNSNNHNSKDISNRLTNDLSQELARLKNLNEGMYFTATPSASLEVI 108
DB 71 OSQYIVIHNQD-----NLYDDLQFLVKIRD-----IYTKLGSVAIEIL 110

RESULT 5
US-10-037-417-59
; Sequence 59, Application US/10037417
; Patent No. 6903201
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Alsobrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Patturajan, Meera
; APPLICANT: Grosse, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Sureeh G
; APPLICANT: Anderson, David W
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Miller, Charles E
; APPLICANT: Eisen, Andrew J
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-235
; CURRENT APPLICATION NUMBER: US/10/037,417
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/260,018
; PRIOR FILING DATE: 2001-01-05
```

```
; PRIOR APPLICATION NUMBER: 60/260,360
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/272,411
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/272,817
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/291,186
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/303,231
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/305,060
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/318,405
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/318,700
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 227
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 59
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Myosin Tail
; OTHER INFORMATION: Consensus Sequence
US-10-037-417-59

Query Match      13.5%; Score 79.5; DB 2; Length 860;
Best Local Similarity 29.0%; Pred. No. 6.5;
Matches 29; Conservative 13; Mismatches 49; Indels 9; Gaps 3;

Qy 26 QIQDALSSATTSLSALTQSNDRGSGLLSSFLRKQNSN-----NHSKDIQNLRTINDSL 79
Db 184 QLESQSLQVKQLDELQQLNDLTS--QKSRLQSENLDLTRLQLEAEAAQVSNLSKLSQ 241
Qy 80 SQELARLKSNNEGMFTATPSASLEVIKYDLSYLKRL 119
Db 242 ESQLEAKRSLEESR--BRANLQAQLRQLEHDLQSLRQ 280

RESULT 6
US-09-248-796A-20654
; Sequence 20654, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20654
; LENGTH: 1290
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (56), (60), (62), (65), (66), (109)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-20654

Query Match      13.5%; Score 79; DB 2; Length 1290;
Best Local Similarity 21.6%; Pred. No. 13;
Matches 29; Conservative 25; Mismatches 46; Indels 34; Gaps 4;

Qy 20 KDESIVQIQDALSSATTS-----LSALTQSNDRGSGLLSSFLRK-----QNSNNHSDI 69
Db 593 EDELTPLEDNLGTSSEFDEKSFMTKLKEADIESDGTMRKSVLRLMKLKQSERSTEDV 652
```

```
Qy 70 SNLRTINDLSQ-----LARKSNLNNEGMFTATP-----SASL 105
Db 653 HNMLVLVDSMLQHVTVNESRVIGTDAHSLVNLITIQKLDRLSTEDMARVAESKMSLSQL 712
Qy 106 EVIKYDLSYLKRL 119
Db 713 ELVKEEKELLEKEL 726
```

RESULT 7

```
US-09-248-796A-19426
; Sequence 19426, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19426
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19426
```

```
Query Match      13.4%; Score 78.5; DB 2; Length 333;
Best Local Similarity 23.6%; Pred. No. 2.2;
Matches 32; Conservative 23; Mismatches 36; Indels 17; Gaps 6;
```

```
Qy 18 FYKDES---IVQ-LQDAL--SSATTSLSALTQSNDRGSGLLSSFLRKQNSNNHSDISN 71
Db 228 FWKDDSATFLVQKLQELTRSEFDIVSSTNNQ-----SSFARDQQAANAIQIK 281
```

```
Qy 72 LRTINDLSQELARLKSNNEGMFTATPSASLEVIKYDLSYLKRL 119
Db 282 LKLIARENSDQSKLPGNPN--MILSVHATATPKSV---FAFLKQL 324
```

RESULT 8

```
US-09-540-236-3533
; Sequence 3533, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3533
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-3533
```

```
Query Match      13.4%; Score 78.5; DB 2; Length 1627;
Best Local Similarity 27.0%; Pred. No. 20;
Matches 24; Conservative 18; Mismatches 38; Indels 9; Gaps 2;
```

```
Qy 31 LSSATTSLSALTQSNDRGSGLLSSFLRKQNSNNHSDISNLRITINDLSQELARLKSNI 90
Db 882 LSSALVIDNTLKQNNAKTQEIAN--LTQTVSGHTSQVRELVTGDLSSQKYSQVKTQA 938
```

```
Qy 91 NNEGMFTAT-----PSASLEVIKYDLS 113
```

Db 939 NNATSEITAKOTGGQATSVIRSELA 967

RESULT 9

US-09-248-796A-20889
; Sequence 20889, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09248796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20889
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20889

Query Match 13.0%; Score 76.5; DB 2; Length 214;
Best Local Similarity 22.5%; Pred. No. 2;
Matches 36; Conservative 21; Mismatches 48; Indels 55; Gaps 6;
Qy 3 IYVEIERRIENHFLFYKDESIVQLQDALSSATTSLSALTQSNDRGS-----GLLSFL 57
Db 5 VYFEIDA-----TDDQLIEYQDRLEIATTTTAMAIQENISEKSPFALIGSDETI 54
Qy 58 RKQNNHNS-KDISNRLTNDLSQ-----ELARK-----SNLN 91
Db 55 QVSLQEHNCLEIKSLQALSANSSSLKMLKLSGKETHVHIGMENGYYARIKIDTINGNLS 114
Qy 92 NEGMYFTATPSASLEVIKYD-----LSYLKR 117
Db 115 NSRVKIIGKPVSLSVIKFNSNEIEGILAISSAPWISLYR 154

RESULT 10

US-09-543-681A-8041
; Sequence 8041, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8041
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8041

Query Match 13.0%; Score 76.5; DB 2; Length 367;
Best Local Similarity 27.4%; Pred. No. 4.2;
Matches 26; Conservative 17; Mismatches 31; Indels 21; Gaps 4;
Qy 35 TTSLGALTQSNDRGSLLSFLRKQNSN---NHGKDISNRLTNDLSLSQELA-RLKSNL 90
Db 11 TNYLSLVTONLNKSGTGLSIAERLSGLRINSKD-----DAACQAIANFTSNV 62
Qy 91 N-----NEGMYFTATPSASLEVIKYDLSYLK 116
Db 63 NGLTOASRNANDGISIAQTTEGALNEINNLRQR 97

RESULT 11

US-09-543-681A-4980
; Sequence 4980, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4980
; LENGTH: 2807
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4980

Query Match 13.0%; Score 76.5; DB 2; Length 2807;
Best Local Similarity 26.7%; Pred. No. 72;
Matches 32; Conservative 26; Mismatches 33; Indels 29; Gaps 8;
Qy 3 IYVEIERRIENHFLFYKDESIVQLQDALSS---ATTLSALTQSNDRGSLLSFLRK 59
Db 982 IYLD-----ENGKFNV-TGKNNLNQDKLISGKNAITSIKRLVE--KEYGSEITNQVFNQ 1033
Qy 60 --QN---SNHSDKDISNRLTNDLSLSQELARKLSNLANEGMFYATPS-----ASLEV 107
Db 1034 FTENEPAGNGHIDLSGLKKHQAEQQVSPISSTL-----YIWKPSKHSRLGHAAQOI 1087

RESULT 12

US-08-260-582-77
; Sequence 77, Application US/08260582
; Patent No. 5635182
; GENERAL INFORMATION:
; APPLICANT: McCoy, John M.
; APPLICANT: Lu, Zhijian
; TITLE OF INVENTION: METHOD OF DETECTING LIGAND
; TITLE OF INVENTION: INTERACTIONS
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/260,582
; FILING DATE: 16-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meinert, M. C.
; REGISTRATION NUMBER: 31,544
; REFERENCE/DOCKET NUMBER: GI 5236
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

Query Match      12.9%; Score 75.5; DB 2; Length 569;
Best Local Similarity 23.8%; Pred. No. 10;
Matches 30; Conservative 25; Mismatches 44; Indels 27; Gaps 4

QY      15 NLFYKIDSLVOLQDALSATTSLSALTQSNDDR-----GSGLL 53
Db      194 NSILLSETSLSSDDSTSS--DTSISSTTSDTGNINAGSLHTSTASIKELSIQKTGVTL 252

QY      54 SSFIRKQNNHNSKDISNLRITLNDLSLQELARLKSNNLNEGMYTATPSASLEVIKYDLS 113
Db      253 SSYLSTKLSS--TSDITIELITTELIITTELTIE--DNEPNTFTTPSSSHSEIFSSDNS 307

QY      114 YLKRL 119
Db      308 VLSKQV 313

RESULT 15
US-07-603-133B-13
; Sequence 13, Application US/07603133B
; Patent No. 5298244
; GENERAL INFORMATION:
; APPLICANT: Redmond, Mark J.
; APPLICANT: Ijaz, Mohammed K.
; APPLICANT: Parker, Michael D.
; TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
; USE IN A VACCINE TO ROTAVIRAL DISEASE
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 545 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: CA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/603,133B
; FILING DATE: 19901025
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9313-0004.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 327-7250
; TELEFAX: (415) 327-2951
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 775 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-603-133B-13

Query Match      12.8%; Score 75; DB 1; Length 775;
Best Local Similarity 27.4%; Pred. No. 17;
Matches 26; Conservative 15; Mismatches 40; Indels 14; Gaps 3

QY      32 SSATTSALTQSNND-----RSGLLSFLRKQNNHNSKDISNLRITLNDLSLQELAR 85
Db      554 SKLATSISEMTHLSLDAASSASRSVSRISNITISNWTNVSDNSVNTLSLSDISTQTST 613

QY      86 LKSNL-----NNEGMYTATPSASLEVIKYDLS 113
Db      614 ISKNRLKEMITQTEGMSFDDIGAAVLKT-KIDMS 647

```

Search completed: January 24, 2006, 20:02:37
Job time : 8.3361 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 25.5705 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-2

Perfect score: 587

Sequence: 1 QKIYVEIERRIENHFLFYK.....TPSASLEVIKYDLSYLRRL 119

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	566	96.4	407	Q9R3K2 BORBU	Q9r3k2 borrelia bu
2	566	96.4	407	Q9S0G7 BORBU	Q9s0g7 borrelia bu
3	560	95.4	407	Q9S0A2 BORBU	Q9s0a2 borrelia bu
4	553	94.2	407	Q5XYN9 BORGA	Q5xyn9 borrelia ga
5	531.5	90.5	408	Q9S0S7 BORBU	Q9s0s7 borrelia bu
6	529	90.1	407	Q5XZ87 BORGA	Q5xz87 borrelia ga
7	508.5	86.6	341	Q5XYV6 BORGA	Q5xyv6 borrelia ga
8	300	51.1	404	Q9KKC1 BORHE	Q9kcc1 borrelia he
9	264.5	45.1	413	Q50931 BORBU	Q50931 borrelia bu
10	243	41.4	223	Q6ASJ2 BORGA	Q6asj2 borrelia ga
11	105	17.9	140	Q5XZ76 BORGA	Q5xz76 borrelia ga
12	94	16.0	581	Q7RQ99 PLAYO	Q7rq99 plasmodium
13	93	15.8	525	Q81IE7 PLAF7	Q81ie7 plasmodium
14	92.5	15.8	1056	Q7W111 HELHP	Q7w111 helicobacte
15	92	15.7	506	Q5SE55 DICDI	Q5se55 dictyosteli
16	91.5	15.6	1090	Q54Y90 DICDI	Q54y90 dictyosteli
17	91	15.5	326	Q9J1B3 GREOV	Q9j1b3 avian reovi
18	91	15.5	715	Q54G03 DICDI	Q54g03 dictyosteli
19	90.5	15.4	407	Q9WZHO THEMA	Q9wzho thermotoga
20	90.5	15.4	621	Q54PZ7 DICDI	Q54pz7 dictyosteli
21	90.5	15.4	2910	Q81BY8 PLAF7	Q81by8 plasmodium
22	90	15.3	646	Q6MTK5 MYCMA	Q6mtk5 mycoplasma
23	89.5	15.2	936	Q801N8 XENLA	Q801n8 xenopus lae
24	89.5	15.2	954	Q55FSA DICDI	Q55fsa dictyosteli
25	89.5	15.2	3130	Q9BK46 PLAF7	Q9bk46 plasmodium
26	89.5	15.2	3130	Q81DX6 PLAF7	Q81dx6 plasmodium
27	89.5	15.2	3254	Q7YWE9 PLAF7	Q7ywe9 plasmodium
28	89.5	15.2	3254	Q9BK45 PLAF7	Q9bk45 plasmodium
29	89	15.2	648	Q54G04 DICDI	Q54g04 dictyosteli
30	89	15.2	3080	Q7YWF0 PLAF7	Q7ywf0 plasmodium
31	89	15.2	3081	Q7YWF1 PLAF7	Q7ywf1 plasmodium

32 89 15.2 3203 2 Q7YWE6 PLAF7
33 89 15.2 3203 2 Q7YWE7 PLAF7
34 88.5 15.1 1141 2 Q55EG1 DICDI
35 88.5 15.1 3096 2 Q7YWF2 PLAF7
36 88.5 15.1 3256 2 Q7YWE8 PLAF7
37 88 15.0 503 2 Q15655 HUMAN
38 88 15.0 687 2 Q58P23 HUMAN
39 88 15.0 692 2 Q504U6 HUMAN
40 88 15.0 726 2 Q15624 HUMAN
41 88 15.0 1681 2 Q81J39 PLAF7
42 88 15.0 2328 2 Q4PGJ7 USTMA
43 88 15.0 2349 1 TPR HUMAN
44 88 15.0 2363 2 Q99968 HUMAN
45 88 15.0 2363 2 Q5SWY0 HUMAN

ALIGNMENTS

RESULT 1
Q9R3K2 BORBU PRELIMINARY; PRT; 407 AA.
AC Q9R3K2
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE Hypothetical protein BBL01 (Hypothetical protein BBS01) (Hypothetical protein BBS01) (Hypothetical protein BBS01)
GN OrderedLocNames=BBL01, BBS01, BBS01
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-8, Plasmid cp32-3, Plasmid cp32-6, and Plasmid cp32-1.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_taxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=cp32-8, cp32-3, cp32-6, and cp32-1;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P., Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H., Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RA "A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=cp32-8, cp32-3, cp32-6, and cp32-1;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K., Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D., Salzberg S.L., Peterson J.D., Kerlavage A.R., Quackenbush J., Richardson S.L., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Bowman C., Garland S.A., Fujii C., Cotton M.D., L.A., Artiach P., Hach B., Smith H.O., Venter J.C.;
RA "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AE001580; AAF07646.1; -; Genomic_DNA.
DR EMBL; AE001576; AAF07474.1; -; Genomic_DNA.
DR EMBL; AE001578; AAF07561.1; -; Genomic_DNA.
DR EMBL; AE001575; AAF07429.1; -; Genomic_DNA.
DR TIGR; BBL01; -;
DR TIGR; BBS01; -;
DR TIGR; BBS01; -;
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 407 AA; 47037 MW; 6440A2466F5B9F23 CRC64;

RESULT 3					
Q9SOA2_BORBU					
ID	Q9SOA2 BORBU PRELIMINARY;	PRT;	407 AA.		
AC	Q9SOA2;				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	Hypothetical protein BBO01.				
GN	OrderedLocuaNames=BBO01;				
OS	Borrelia burgdorferi (Lyme disease spirochete) .				
OC	Plasmid cp32-7.				
OC	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;				
OC	Borrelia burgdorferi group.				
OX	NCBI_TaxID=139;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=ATCC 35210 / B31;				
RC	MEDLINE=20138354; PubMed=10672174;				
RA	Casjens S., Palmer N.M., van Vugt R., Huang W.M., Stevenson B., Rosa P.,				
RA	Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,				
RA	Hickey E.K., Gwinn M.L., White O., Fraser C.M.;				
RT	"A bacterial genome in flux: the twelve linear and nine circular				
RT	extrachromosomal DNAs in an infectious isolate of the Lyme disease				
RT	spirochete Borrelia burgdorferi."				
RL	Mol. Microbiol. 35:490-516(2000).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=ATCC 35210 / B31;				
RX	MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;				
RA	Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,				
RA	Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,				
RA	Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,				
RA	Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,				
RA	Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,				
RA	Gocayne J.D., Weidman J.F., Uitterback T.R., Watthey L., McDonald L.A.,				
RA	Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,				
RA	Roberts K.M., Hatch B., Smith H.O., Venter J.C.;				
RT	"Genomic sequence of a Lyme disease spirochete, Borrelia				
RT	burgdorferi."				
RL	Nature 390:560-586(1997).				
DR	EMBL; AE001579; AAF07602.1; -; Genomic_DNA.				
DR	TIGR; BBO01; -;				
DR	InterPro; IPR009427; DUF1073.				
DR	Pfam; PF06381; DUF1073; 1.				
KW	Complete proteome; Hypothetical protein; Plasmid.				
SQ	SEQUENCE 407 AA; 47061 MW; 8E34A6BF392D15EE CRC64;				
Query Match		95.4%;	Score 560;	DB 2;	Length 407;
Best Local Similarity		94.1%;	Fred. No. 6.8e-39;		
Matches 112;		Conservative 4;	Mismatches 3;	Indels 0;	Gaps 0;
Qy	1 QKIVETRIENHFLFYKDLSVLQDALSSATTSLTSALTQSNDRGSGILSSPLRKQ 60				
	:				
Dd	174 EKIVETRIENHFLFYKDLSVLQDALSSATTSLTSALTQGNDRGSGILSSPLRKQ 233				
	:				
Qy	61 NSNNHKDINSIRLTINSLSQELARKLNINNEGMPYTTATPSASLEVIKYDL SylKRL 119				
	:				
Dd	234 NSNNHKDINSIRLTINSLQAQLARKLNINNEGMPYTTATPSASLEVIKYDL SylKEAL 292				
	:				

RESULT 4	
Q5XYN9 BORG	
ID Q5XYN9 BORG PRELIMINARY;	PRT; 407 AA.
AC Q5XYN9	
DT 25-OCT-2004 (TREMBLrel. 28, Created)	
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)	
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)	
DE Hypothetical protein.	
GN ORENAMES-BGP242;	
OS Borrelia garinii PBI.	
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;	
OC Borrelia burgdorferi group.	
NCBI TaxID=290434;	
OX	

```

RN NUCLEOTIDE SEQUENCE.
RP STRAIN=PB1;
RA Glockner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Schilhabel M., Wilske B., Suhnel J., Platzter M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Nucleic Acids Res. 32:6038-6046(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PB1;
RA Glockner G., Schilhabel M., Lehmann R., Platzter M.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY722930; AU86093.1; -; Genomic_DNA.
DR InterPro: IPR009427; DUF1073.
DR Pfam: PF06381; DUF1073; 1.
KW Hypothetical protein.
SQ SEQUENCE 407 AA; 46848 MW; 8C13E7AE0350AE79 CRC64;

Query Match 94.2%; Score 553; DB 2; Length 407;
Best Local Similarity 93.3%; Pred. No. 2.6e-38;
Matches 111; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 QKIYVEIERRIENHFLFYKDESLVQLQDALSSATTSLTSALTQSNDRGSGILSSFLRKQ 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 EKIVVEIERRIENHFLFYKDESLVQLQDALSSATTSLTSALTQSNDRGSGILSSFLRKQ 233

QY 61 NSNNHSDISNLRITLNDLSQELARLKSNNLNNEGMFYTATPSASLEVIKYDLSYLKRL 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 NSNNHSDISNLRITLNDLSQELARLKSNNLNNEGMFYTATPSASLEVIKYDLSYLKRL 292

RESULT 5
Q9S057 BORBU
ID Q9S057 BORBU PRELIMINARY; PRT; 408 AA.
AC Q9S057;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BBN01.
GN OrderedLocusNames=BBN01;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Castjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Castjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Wattley L., McDonald L.A.,
RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL: AE001581; AAF07681.1; -; Genomic_DNA.
DR TIGR; BBN01; -;

DR InterPro: IPR009427; DUF1073.
DR Pfam: PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 408 AA; 47038 MW; 5EB60167E175F272 CRC64;

Query Match 90.1%; Score 529; DB 2; Length 407;
Best Local Similarity 87.4%; Pred. No. 2.8e-36;
Matches 104; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 1 QKIYVEIERRIENHFLFYKDESLVQLQDALSSATTSLTSALTQSNDRGSGILSSFLRKQ 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174 EKIVVEIERRIENHFLFYKDESLVQLQDALSSATTSLTSALTQSNDRGSGILSSFLRKQ 233

QY 61 NSNNHSDISNLRITLNDLSQELARLKSNNLNNEGMFYTATPSASLEVIKYDLSYLKRL 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 NSNNHSDISNLRITLNDLSQELARLKSNNLNNEGMFYTATPSASLEVIKYDLSYLKRL 292

RESULT 7
Q5XYI6 BORGA
ID Q5XYI6 BORGA PRELIMINARY; PRT; 341 AA.
AC Q5XYI6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BGPI45;
OS Borrelia garinii PB1.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
```

```

OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PB1;
RA Glockner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Schillhabel M., Wilske B., Suhnel J., Platzner M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Nucleic Acids Res. 32:6038-6046(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PB1;
RA Gloeckner G., Schillhabel M., Lehmann R., Platzner M.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY729222; AAU85996.1; -; Genomic_DNA.
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Hypothetical protein.
SQ SEQUENCE 341 AA; 39289 MW; EF138B65AD40F827 CRC64;

Query Match      86.6%; Score 508.5; DB 2; Length 341;
Best Local Similarity 85.0%; Pred. No. 1.2e-34;
Matches 102; Conservative 10; Mismatches 7; Indels 1; Gaps 1;

QY 1 QKIYVEIERRIENHNFLFYKDESLVQLQDALSSATTSLSALTQ-SNNDRGSGLLSFLRK 59
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 EKIYVEIEKRIENHNFLFYKDESLVQLQDALSSATTSLSLTQNSSEKSGILLSFLRK 235
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 60 QNSNNHSDISNRLTNDLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKRL 119
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 236 QNSNNHSDLSNRLSNDLSLAHELLKNNLNNEGMYTANPSASLEVIKYDLSYLKEVL 295

RESULT 8
Q9KKC1 BORHE
ID Q9KKC1 BORHE PRELIMINARY; PRT; 404 AA.
AC Q9KKC1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Borrelia hermsii.
OG Plasmid cp32.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-1;
RX MEDLINE=20316012; PubMed=10858201;
DOI=10.1128/IAI.68.7.3900-3908.2000;
RA Stevenson B., Porcella S.F., Oie K.L., Fitzpatrick C.A., Raffel S.J.,
RA Lubke L., Schrumpf M.E., Schwan T.G.;
RT "The relapsing fever spirochete Borrelia hermsii contains multiple,
RT antigen-encoding circular plasmids that are homologous to the cp32
RT plasmids of Lyme disease spirochetes.";
RL Infect. Immun. 68:3900-3908(2000).
DR EMBL: AF123078; AAF28886.1; -; Genomic_DNA.
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 404 AA; 46409 MW; 5567119DBB024671 CRC64;

Query Match      51.1%; Score 300; DB 2; Length 404;
Best Local Similarity 50.8%; Pred. No. 5.1e-17;
Matches 62; Conservative 24; Mismatches 28; Indels 8; Gaps 2;

QY 1 QKIYVEIERRIENHNFLFYKDESLVQLQDALSSATTSLSALTQSNNDRGSGLLSFLRK 57
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 170 EKIYVEIEKRIENHNFLFYKDESLVQLQDALSSATTSLSLTQNSSEKSGILLSFLRK 229
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 58 RKQNSNNHSDISNRLTNDLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLK 117
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 SKEDG-----DVSSLSRANDELMRELERLKSLLDNDGIFYSCEPNASLEVIKYDLSYLKE 284

```

```

QY 118 RL 119
Db 285 AL 286

RESULT 9
O50931 BORBU
ID O50931 BORBU PRELIMINARY; PRT; 413 AA.
AC O50931;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BBA38.
GN OrderedLocusNames=BBA38;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-P., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL: AE000790; AAC66300.1; -; Genomic_DNA.
DR PIR; F70211; F70211.
DR TIGR; BBA38; -.
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 413 AA; 47901 MW; 382654B2DF31241 CRC64;

Query Match      45.1%; Score 264.5; DB 2; Length 413;
Best Local Similarity 47.4%; Pred. No. 5.1e-14;
Matches 55; Conservative 26; Mismatches 28; Indels 7; Gaps 3;

QY 1 QKIYVEIERRIENHNFLFYKDESLVQLQDALSSATTSLSALTQSNNDRGSGLLSFLRK 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 EQIYLEIEKRIENHNFLFYKDEHLVGLVLESLAKEEINVLA--NSKGK-IFSTFFKQA 241
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 NSNNHSDISNRLTNDLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLK 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 242 EPN---KSFQALSSVSDLSKLNLTNDGIFYTASENARLEVIKYDLEFLK 294
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
Q6ASJ2 BORG
ID Q6ASJ2 BORG PRELIMINARY; PRT; 223 AA.
AC Q6ASJ2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BGA33;
OS Borrelia garinii.
OG Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PB1;

```

```

RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RT Schulte-Spechtel U., Wilske B., Suhnel J., Platzter M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: CP000015; AAT93792.1; -; Genomic_DNA.
DR InterPro: IPR009427; DUF1073.
DR Pfam: PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 223 AA; 25591 MW; E210E27A83C386D4 CRC64;

Query Match 41.4%; Score 243; DB 2; Length 223;
Best Local Similarity 49.1%; Pred. No. 1.6e-12;
Matches 54; Conservative 21; Mismatches 29; Indels 6; Gaps 3;

QY 7 IERRIENHNFYKDESLVQLQDLSATTSALTSQNNDRGSLLSFLRKQNSNHS 66
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MEKIRNYNFFPYKDEHLVGLTESLELAKEINVLV--NNNGK-IFSSFFKGQPN-- 54

QY 67 KDISNLTNDLSDELARLKSNNNEGMFYTATPSASLEVTKYDLSYLK 116
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 KSFSALSSASSKSLISLKSIKSLTNNDGIFYTASENASPEVIKYDLEFLK 104

RESULT 11
QXZ76_BORGA
ID Q5XZ76_BORGA PRELIMINARY; PRT; 140 AA.
AC Q5XZ76;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BG056;
OS Borrelia garinii PBI.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RT Schulte-Spechtel U., Schilhabel M., Wilske B., Suhnel J., Platzter M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Nucleic Acids Res. 32:6038-6046(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Schilhabel M., Lehmann R., Platzter M.;
RA Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY72919; AAU85906.1; -; Genomic_DNA.
DR InterPro: IPR009427; DUF1073.
DR Pfam: PF06381; DUF1073; 1.
KW Hypothetical protein.
SQ SEQUENCE 140 AA; 15945 MW; 4B9DD0E998A523DF CRC64;

Query Match 17.9%; Score 105; DB 2; Length 140;
Best Local Similarity 88.0%; Pred. No. 0.38;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 95 MFYTATPSASLEVTKYDLSYLRRL 119
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MFYTATPSASLEVTKYDLSYLRRL 25

RESULT 12
QYQ99_PLAYO
ID QYQ99_PLAYO PRELIMINARY; PRT; 581 AA.
AC QYQ99;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY01203;
OS Plasmodium yoelii yoelii.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguoli S.V., Suh B.B., Kooij T.W., Perte M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01000316; EAA20493.1; -; Genomic_DNA.
DR InterPro: IPR007275; YTH.
DR Pfam: PF04146; YTH; 1.
DR PROSITE: PS50882; YTH; 1.
KW Hypothetical protein.
SQ SEQUENCE 581 AA; 68268 MW; B41B699C83A6B986 CRC64;

Query Match 16.0%; Score 94; DB 2; Length 581;
Best Local Similarity 28.7%; Pred. No. 17;
Matches 31; Conservative 18; Mismatches 41; Indels 18; Gaps 5;

QY 2 KIYVE--IERRIENHNFYKDESLVQLQDLSATTSLSA-----LTQSNDRSG 51
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
213 KFYLDQILINKNYTH---LYKSQVLNDMKNKLDALIVLSEKYEKNISLTKERENN 269

QY 52 LLSSFLRKQNSNNHKSIDISNLTNDLSQELARLKSNNNEGMFYTA 99
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 TLSDMF--TNSDNNYKDISNRLNLSGIN---IHDDKNLNRSGIINTA 312

RESULT 13
Q8IIE7_PLAF7
ID Q8IIE7_PLAF7 PRELIMINARY; PRT; 525 AA.
AC Q8IIE7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF11_0227;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallom S.J., Suh B., Peterson J., Anguoli S.,
RA Perte M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McRadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum";
RL Nature 419:498-511(2002).
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL: AE014839; AAN35811.1; -; Genomic_DNA.
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.

```

Search completed: January 24, 2006, 19:56:39
Job time : 28.5705 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 4.51452 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-2
Perfect score: 587
Sequence: 1 QKIYVEIERRIENHFLYK.....TPSASLEVIKYDLSYLKRL 119

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264.5	45.1	413	2 F70211	hypothetical prote
2	90.5	15.4	407	2 F72343	hypothetical prote
3	88	15.0	503	4 S23741	hypothetical TPR/T
4	88	15.0	2094	2 S33124	tpx protein - huma
5	85.5	14.6	3394	2 T18501	hypothetical prote
6	85	14.5	348	2 I36839	C9L protein - vari
7	84	14.3	602	1 TVRTR	protein kinase (SC
8	83	14.1	736	1 VPXRPC	outer layer protei
9	83	14.1	1091	2 T34107	hypothetical prote
10	82	14.0	2061	2 T13751	transcription fact
11	81	13.8	309	1 S34198	IgE Fc receptor II
12	81	13.8	561	2 B90414	hypothetical prote
13	81	13.8	1474	2 T18281	hypothetical prote
14	80.5	13.7	666	2 S50452	hypothetical prote
15	80.5	13.7	790	2 F75372	c-di-GMP phosphodi
16	80	13.6	321	2 A42507	F5L protein - vacc
17	80	13.6	322	2 G36213	F5L protein - vacc
18	80	13.6	1302	1 JC6009	surface-located me
19	79.5	13.5	2712	2 T05113	hypothetical prote
20	79	13.5	348	2 T28467	major envelope ant
21	79	13.5	348	2 C72154	E5L protein - vari
22	79	13.5	415	2 JQ1212	Mob protein - Bac
23	79	13.5	415	2 S05987	plasmid recombinat
24	79	13.5	1051	2 T18351	lmpI protein - Myc
25	79	13.5	1365	2 T30822	hypothetical prote
26	79	13.5	2573	2 D71614	hypothetical prote
27	78.5	13.4	496	2 D97728	thermostable carbo
28	78.5	13.4	950	2 E70203	exonuclease sbcc (
29	78	13.3	1036	2 B69368	hypothetical prote

ALIGNMENTS

RESULT 1

F70211
hypothetical protein BBA38 - Lyme disease spirochete plasmid A/lp54
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: F70211
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70211
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-413 <KLE>
A:Cross-references: UNIPROT:O50931; UNIPARC:UPI000005691F; GB:AE000790; NID:G2690224; P1
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid
C:Superfamily: Borrelia burgdorferi hypothetical protein BBA38

Query Match 45.1%; Score 264.5; DB 2; Length 413;
Best Local Similarity 47.4%; Pred. No. 1.4e-15;
Matches 55; Conservative 26; Mismatches 28; Indels 7; Gaps 3;

QY 1 QKIYVEIERRIENHFLYKDESLVQLQDALSSATTSLSALTQSNDRGSGLLSFLRKQ 60
Db 186 EQVLEIEKRINRYNFLFYKDEHLVGLVESLEIAKEINVLA--NSKGK-IFSTFFKRAQ 241
QY 61 NSNNHSDKIDISNLRITANDLSQELARLKNANNEGMYTATPSASLEVIKYDLSYLK 116
Db 242 EPN---KSFQALSSVSDLSRELSKIKTNLNDGIFYTASENARLEVIKYDLEFLK 294

RESULT 2

F72343
hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: F72343
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: F72343
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <ARN>

Db 1854 YIELKRLKTEKTKINQEKYIILOQKDNLI-LDNFNSTTT-----TNNNN----- 1901

Qy 54 SSFLRKQNSNNHSDISNRLTNDLSQELARLKSNN 92

Db 1902 -----NDNNNDNNNDNNNDNYQQFIHSLKANLEN 1933

RESULT 6

C9L protein - variola virus (strain India-1967)

C:Species: variola virus

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004

C:Accession: I36839

R:Blinov, V.M.

submitted to GenBank, November 1992

A:Reference number: A36859

A:Accession: I36839

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-348 <BLI>

A:Cross-references: UNIPROT:P33865; UNIPARC:UPI0000167901; GB:X69198; NID:g456758; PIDN:

C:Superfamily: phospholipase D, chordopox type

Query Match 14.5%; Score 85; DB 2; Length 348;

Best Local Similarity 24.8%; Pred. No. 4.7;

Matches 27; Conservative 22; Mismatches 34; Indels 26; Gaps 3;

Qy 13 NHNLFYKDESILVQLQDALSSATTSLSALTSQNDNRGSLSSFLRKQNSNNHSDISN 72

Db 67 NHTFWYNNK-----VIALATEDRTSGYISFFIKRWISLTCINISL 109

Qy 73 RTLNDLSQELARLKSNNHSGMPYATPSASLEVIKYDLS-----YLKR 117

Db 110 RYEDSGSYKGVSHLK-----DGVIVTTTMSVANKIIDLGRVCYLTR 153

..

RESULT 7

TVTRNR

protein kinase (EC 2.7.1.37) raf - rat

N:Alternate names: kinase-related transforming protein raf; raf proto-oncogene protein-8

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 31-Dec-2004

C:Accession: B26126

R:Ishikawa, F.; Takaku, F.; Nagao, M.; Sugimura, T.

Mol. Cell. Biol. 7, 1226-1232, 1987

A:Title: Rat c-raf oncogene activation by a rearrangement that produces a fused protein.

A:Reference number: A26126; MUID:87172791; PMID:3550433

A:Accession: B26126

A:Molecule type: mRNA

A:Residues: 1-602 <ISH>

A:Cross-references: UNIPARC:UPI000011P8CA; GB:M15428; NID:g206546; PIDN:AAA42002.1; PID:

C:Genetics:

A:Gene: raf

C:Keywords: ATP; autophosphorylation; phosphoprotein; phosphotransferase; serine/threonine

F:301-567/Domain: protein kinase homology <KIN>

F:309-317/Region: protein kinase ATP-binding motif

F:329/Active site: Lys #status predicted

F:453/Binding site: phosphate (Ser) (covalent) (by autophosphorylation) #status predicted

Query Match 14.3%; Score 84; DB 1; Length 602;

Best Local Similarity 26.9%; Pred. No. 11;

Matches 36; Conservative 24; Mismatches 52; Indels 22; Gaps 8;

Qy 4 YVEIERRI-ENHNFYKDESILVQLQDALSSATTSLSALTQSN-----NDRGSLSSP 56

Db 54 YFEIEKLSQSOERLVNRETRECONLRLEKLNQVKVLTENKKELETAQDENLGIQSOF 113

Qy 57 LR-KQNSNNHSDISNRLTNDLSQELARLKSNN--NEGMYFATPSASLEVIKYD-- 111

Db 114 TRAKEELAEKRD--IRT-NERLSQEVYLTEDVKRLNEKLKESNTTKGELQ-LKLDL 169

Qy 112 -----LSYLKERL 119

Db 170 QASDVTVKYREKRL 183

RESULT 8

VPXRPC

outer layer protein VP4 - porcine rotavirus C (strain Cowden)

N:Alternate names: hemagglutinin; outer capsid protein VP4

N:Contains: outer capsid protein VP5; outer capsid protein VP8

C:Species: porcine rotavirus C

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 09-Jul-2004

C:Accession: D40822

R:Bremont, M.; Juste-Lesage, P.; Chabanne-Vautherot, D.; Charpilienne, A.; Cohen, J.

Virology 186, 684-692, 1992

A:Title: Sequences of the four larger proteins of a porcine group C rotavirus and comparison with other group C rotaviruses

A:Reference number: A40822; MUID:92124743; PMID:1310192

A:Accession: D40822

A:Molecule type: genomic RNA

A:Residues: 1-736 <BRE>

A:Cross-references: UNIPROT:P26193; UNIPARC:UPI0000138C59; GB:M74218; NID:g333317; PIDN:

C:Genetics:

A:Map position: segment 4

C:Superfamily: rotavirus outer layer protein VP3

C:Keywords: Glycoprotein; hemagglutinin; outer capsid protein

F:1-243/Product: outer capsid protein VP8 #status predicted <VP8>

F:244-250/Region: cleavage processing #status predicted

F:251-736/Product: outer capsid protein VP5 #status predicted <VP5>

F:39,61,64,93,162,191,237,251,304,471,631/Binding site: carbohydrate (Asn) (covalent) #

Query Match 14.1%; Score 83; DB 1; Length 736;

Best Local Similarity 22.2%; Pred. No. 18;

Matches 38; Conservative 29; Mismatches 46; Indels 58; Gaps 7;

Qy 3 IYVEIERRIENHNFYKDESILVQLQDALSSATTT-----SLSALTQSNDRGSL 53

Db 494 VRMDLERKINDLRNDFNLANSVALGILSLATSLTANLLESVPALASSVKDVAANVM 553

Qy 54 SSF-----LRKQNSNNH-----KDLSN-----LRTLN----- 76

Db 554 KKFRTNTRFKATKAKYSEFIIGLLEDVTNVARNNGMNFDDITSVMVSTTKLQ 613

Qy 77 -DSLQELARLKSNN-----LNNEGMYFATPSASLEVIKYDLSYLKR 118

Db 614 VDTLSEIVARSADNFIPNRSVRMIEDGIVYEATPK---RTFSYDLTTLQOR 661

RESULT 9

T34107

hypothetical protein C18C4.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34107

R:Gattung, S.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid C18C4.

A:Reference number: Z21478

A:Accession: T34107

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1091 <GAT>

A:Cross-references: UNIPARC:UPI000017B7A1; EMBL:U55369; PIDN:AAC25825.1; GSPDB:GN000023;

A:Experimental source: strain Bristol N2; clone C18C4

C:Genetics:

A:Gene: C18C4.5

A:Map position: 5

A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2; 99

Query Match 14.1%; Score 83; DB 2; Length 1091;

Best Local Similarity 22.9%; Pred. No. 29;

Matches 36; Conservative 22; Mismatches 45; Indels 54; Gaps 6;

Qy 5 VVEIERRIENH-----FLFYKDESILVQLQDAL-----SSATTSLS 39

Db 971 KKKNDQINAELET--KNFKDFNLLNSL---KISNEHKSQNDLNTKNYSLEKEIESLR 1025

QY 88 SNLNNEGMYTATPSASLEVIK--YDSY-----LKR 117

Db 1026 SRI-----IQLETTPTVSNQITQPAPEYSYKHEILKR 1057

RESULT 14

S50452

hypothetical protein YEL007w - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004

C:Accession: S50452

R:Dictrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: Saccharomyces cerevisiae chromosome V cosmids 9871, 8199, 9867, 9495 and

A:Reference number: S50428

A:Accession: S50452

A:Molecule type: DNA

A:Residues: 1-666 <DIE>

A:Cross-references: UNIPROT:P40002; UNIPARC:UPI000013AB94; EMBL:U18530; NID:G602367; PID

C:Genetics:

A:Gene: SGD:TOS9; MIPS:YEL007w

A:Cross-references: SGD:S0000733

A:Map position: 5L

Query Match 13.7%; Score 80.5; DB 2; Length 666;

Best Local Similarity 27.0%; Pred. No. 26;

Matches 24; Conservative 19; Mismatches 37; Indels 9; Gaps 3;

QY 13 NHNFLFYKDES----LVQLQDALSATYLSA----LTQS--NDRGSGLLSFLRKQNSN 63

Db 281 NKNYTYKNDSSGYITQFAPALPSTTLMTYNTANPPYITQSPDNTNATGNTHTVNNNNNS 340

QY 64 NHSKDISNLRTLNDLSQELARLKSNN 92

Db 341 NNSNSNNNNNNNNNNNNNNNNNNNNNN 369

RESULT 15

F75372

c-di-GMP phosphodiesterase A - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: F75372

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: F75372

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-790 <WHI>

A:Cross-references: UNIPROT:Q9RTX3; UNIPARC:UPI00000D3E9E; GB:AE002007; GB:AE000513; NID

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1629

A:Map position: 1

Query Match 13.7%; Score 80.5; DB 2; Length 790;

Best Local Similarity 22.6%; Pred. No. 32;

Matches 28; Conservative 22; Mismatches 59; Indels 15; Gaps 2;

QY 4 YVEIERRIENHNFYKDESILVOLQDALSATYLSALTQSNDRGSGLLSFLRKQNSN 63

Db 174 YAKLEQYDTAQPYLFEAHDHARQLQGFQFDMITSLNLACVQLSGGTGAALTLI--EEAL 231

QY 64 NHSKDISNLRTLN-----DSLSQELARLKSNNLNNEGMYTATPSASLEVIKY 110

Db 232 RHAKTVASPRYKLSVLFHFSQLLDLRLSEALTYCEQGLNTDSGFGTATPQCELVTHG 291

QY 111 DLSY 114

Db 292 BILY 295

Search completed: January 24, 2006, 19:58:58

Job time : 6.51452 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 26.5934 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-2

Perfect score: 587

Sequence: 1 QKIVVEIRRIENHNFYK.....TPSASLEVIKDYLSLKRRLL 119

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	587	100.0	119	8	ADU98743
2	566	96.4	407	8	ADU98745
3	91	15.5	326	2	AY06109
4	88	15.0	503	4	ABG20277
5	88	15.0	726	6	ABP96033
6	88	15.0	2349	6	AAO16359
7	88	15.0	2400	4	ABG20278
8	85.5	14.6	523	4	ABG20276
9	84	14.3	1309	4	ABG30366
10	84	14.3	1984	4	ABE61060
11	83	14.1	650	8	ADN80539
12	83	14.1	1091	8	ADN23675
13	82.5	14.1	761	5	ABB77431
14	82.5	14.1	1396	8	ADN01130
15	82	14.0	2415	4	ABG20279
16	81	13.8	4624	7	ABE78572
17	81	13.8	4624	7	ADU71185
18	80	13.6	186	5	ABP39739
19	80	13.6	186	8	ADS07575
20	80	13.6	1141	7	ABM89610
21	79.5	13.5	336	5	ABP78804
22	79.5	13.5	356	6	ABM68241
23	79.5	13.5	860	5	AD117302
24	79.5	13.5	860	7	ADC38517

25 79 13.5 570 8 ADX94570
26 79 13.5 2573 3 AAB18234
27 79 13.5 4080 8 ADO42289
28 78.5 13.4 350 8 ADY24418
29 78.5 13.4 950 6 ABU19361
30 78.5 13.4 1043 4 ABB71512
31 78.5 13.4 1627 8 ADL05847
32 78.5 13.4 1819 5 AAM47588
33 78 13.3 589 8 ADY12389
34 78 13.3 4498 4 ABB58595
35 77.5 13.2 388 3 AAB21232
36 77.5 13.2 918 4 ABB59892
37 77.5 13.2 1181 8 ADV89275
38 77.5 13.2 1181 8 ADV80528
39 77.5 13.2 1246 4 ABB64281
40 77.5 13.2 1250 3 AAY91279
41 77.5 13.2 1252 8 ADV82565
42 76.5 13.0 367 7 ADF07756
43 76.5 13.0 488 9 ADZ77216
44 76.5 13.0 550 9 AEB91490
45 76.5 13.0 565 4 AAB49642

ALIGNMENTS

RESULT 1
ADU98743
ID ADU98743 standard; protein; 119 AA.
XX
AC ADU98743;
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 2.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
PN WO2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
DR WPI; 2004-834155/82.
XX
DR N-PSDB; ADU98742.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises providing to the subject at least one Borrelia antigen or its fragment.
XX
XX Claim 27; SEQ ID NO 2; 121pp; English.
The invention describes a method of immunizing a subject comprising providing to the subject at least one Borrelia antigen or its fragment to induce an immune response. Also described are: an isolated polynucleotide comprising a sequence having at least 17 contiguous nucleotides in common with a sequence not given in the specification; an isolated polypeptide having at least 5 consecutive amino acids of the sequence not given in the specification; a vaccine composition comprising at least one Borrelia antigen or at least one polynucleotide encoding a Borrelia antigen; screening for at least one test polypeptide or test polynucleotide encoding a polypeptide for an ability to produce an immune response; preparing a vaccine; vaccinating a subject; treating a subject infected


```

Db      85  TTTTLQDSVSTLSTTVTLDTNT-----SSVHSEALSLRTIVDGNSTT 127
QY      83  LARLKNLNNEGMEFYATPSP--ASLE-VIKYDLSY 114
Db      128  IDNLSKSDVSSNGLAITDLSQSRVKSLSVSSHGLSP 162

RESULT 4
ABG20277
ID  ABG20277 standard; protein; 503 AA.
XX
AC  ABG20277;
DT  18-FEB-2002 (first entry)
XX
DE  Novel human diagnostic protein #20268.
XX
KW  Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS  Homo sapiens.
XX
PN  WO200175067-A2.
XX
PD  11-OCT-2001.
XX
PF  30-MAR-2001; 2001WO-US008631.
XX
PR  31-MAR-2000; 2000US-00540217.
PR  23-AUG-2000; 2000US-00649167.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
DR  N-PSDB; AAS84464.
XX
XX  New isolated polynucleotide and encoded polypeptides, useful in
PT  diagnostics, forensics, gene mapping, identification of mutations
PT  responsible for genetic disorders or other traits and to assess
PT  biodiversity.
XX
PS  Claim 20; SEQ ID NO 50636; 103pp; English.
XX
XX  The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC  sequences. (I) is useful as hybridisation probes, polymerase chain
CC  reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC  and in recombinant production of (II). The polynucleotides are also used
CC  in diagnostics as expressed sequence tags for identifying expressed
CC  genes. (I) is useful in gene therapy techniques to restore normal
CC  activity of (II) or to treat disease states involving (II). (II) is
CC  useful for generating antibodies against it, detecting or quantitating a
CC  polypeptide in tissue, as molecular weight markers and as a food
CC  supplement. (II) and its binding partners are useful in medical imaging
CC  of sites expressing (II). (I) and (II) are useful for treating disorders
CC  involving aberrant protein expression or biological activities. The
CC  polypeptide and polynucleotide sequences have applications in
CC  diagnostics, forensics, gene mapping, identification of mutations
CC  responsible for genetic disorders or other traits to assess biodiversity
CC  and to produce other types of data and products dependent on DNA and
CC  amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC  amino acid sequences of the invention. Note: The sequence data for this
CC  patent did not appear in the printed specification, but was obtained in
CC  electronic format directly from WIPO at
CC  ftp.wipo.int/pub/published_pct_sequences
XX
SQ  Sequence 503 AA;

Query Match          15.0%; Score 88; DB 4; Length 503;
Best Local Similarity 29.3%; Pred. No. 2.9;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;

XX  4  YVEIERRIENHNFLPYKDESIV-----QLQDALSSATTSLSALTQSN-----NDRGS 50
Db      54  YFEIEKR-----LSHSQERLVNRETCOSLRLEKLNQKALTEKNKELEIAQDRNI 107
QY      51  GLLSFLR-KQNSNNHSHKDISNRLTNDLSLSQELARLKSNLN--NEGMEFYATPSPASLEV 107
Db      108  AIQSQFTRTKEELEAEKRDLD--IRT-NERLSQELELYLTEDVKRLNEKLKESNTTKGELQ- 163
QY      108  IKYD-----LSYLKRL 119
Db      164  LKLDLQASDVSVKYREKRL 183

RESULT 5
ABP96033
ID  ABP96033 standard; protein; 726 AA.
XX
AC  ABP96033;
XX
DT  01-APR-2000 (first entry)
XX
DE  Human survivin interacting protein TPRI SEQ ID NO:2.
XX
KW  Human; survivin interacting protein; TPRI;
KW  protein histidine phosphatase interacting partner receptor.
XX
OS  Homo sapiens.
XX
PN  WO2003006499-A2.
XX
PD  23-JAN-2003.
XX
PF  11-JUL-2002; 2002WO-EP007729.
XX
PR  13-JUL-2001; 2001EP-00116641.
XX
PA  (MERE ) MERCK PATENT GMBH.
XX
PI  Hentsch B, Hock B;
XX
WPI; 2003-221716/21.
DR  N-PSDB; ABZ76745.
XX
XX  New TPRI polypeptide, useful in diagnostic assays and for identifying
PT  agonist or antagonist compounds of the TPRI polypeptide, which are
PT  potentially useful in therapy.
XX
PS  Claim 2; Page 41-43; 43pp; English.
XX
XX  The present sequence represents human survivin interacting protein TPRI.
CC  TPRI has the ability to bind specifically to protein histidine
CC  phosphatase interacting partner receptor. The TPRI protein can be used in
CC  diagnostic assays and for identifying agonist or antagonist compounds of
CC  the polypeptide TPRI, which are potentially useful in therapy
XX
SQ  Sequence 726 AA;

Query Match          15.0%; Score 88; DB 6; Length 726;
Best Local Similarity 29.3%; Pred. No. 4.8;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;

XX  4  YVEIERRIENHNFLPYKDESIV-----QLQDALSSATTSLSALTQSN-----NDRGS 50
Db      54  YFEIEKR-----LSHSQERLVNRETCOSLRLEKLNQKALTEKNKELEIAQDRNI 107
QY      51  GLLSFLR-KQNSNNHSHKDISNRLTNDLSLSQELARLKSNLN--NEGMEFYATPSPASLEV 107
Db      108  AIQSQFTRTKEELEAEKRDLD--IRT-NERLSQELELYLTEDVKRLNEKLKESNTTKGELQ- 163
QY      108  IKYD-----LSYLKRL 119
Db      164  LKLDLQASDVSVKYREKRL 183
```

```

RESULT 6
AAO16359
ID AAO16359 standard; protein; 2349 AA.
XX
XX AAO16359;
XX
DT 03-APR-2003 (first entry)
XX
DE Human translocated promoter region (TPR) protein, SEQ ID No 7.
XX
KW Human; p53 pathway modulating agent; p53-associated disorder;
translocated promoter region; TPR; cancer.
XX
OS Homo sapiens.
XX
PN WO200299050-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017425.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.
XX
XX (EXEL-) EXELIXIS INC.
XX
XX Friedman L, Florman GD, Belvin M, Francis-Lang H;
XX WPI; 2003-167335/16.
XX
XX Identifying candidate p53 pathway-modulating agents, useful as
PT therapeutic targets for disorders associated with defective p53 function,
PT comprises screening for agents that modulate the activity of translocated
PT promoter region polypeptides.
XX
PS Claim 13; Page 84-94; 94pp; English.
XX
CC The invention comprises a method for identifying a candidate p53 pathway
CC modulating agent. The method involves providing an assay system
CC comprising a purified translocated promoter region (TPR) protein or
CC nucleic acid. The method of the invention is useful for identifying
CC candidate p53 pathway modulating agents for use as therapeutic targets
CC for disorders associated with defective p53 function (e.g. cancer). The
CC TPR proteins and nucleic acids are useful for identifying agents that
CC modulate TPR function. The present amino acid sequence represents a human
CC TPR protein
XX
SQ Sequence 2349 AA;
Query Match 15.0%; Score 88; DB 6; Length 2349;
Best Local Similarity 29.3%; Pred. No. 24;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;
QY 4 YVEIERRIENHFLFYKDESIV-----QLQDALSSATTSLTQSN-----NDRGS 50
DB 54 YFEIEKR-----LSHSQERLVNRETRECQSLRLEKLNQKALTEKRNKELEIAQDRNI 107
QY 51 GLLSSGFLR-KQNSNNHSDISNRLTNDLSQELARLKSNLN--NEGMYFTATPSASLEV 107
DB 108 ATQSQFTTRKELEAEKRLD--IRT-NERLSQELEVLTVKRLNEKLESNTTGGELQ- 163
QY 108 IKYD-----LSYLKRRL 119
DB 164 LKLDLQASDVSVKYREKRL 183
RESULT 7
ABG20278
ID ABG20278 standard; protein; 2400 AA.
XX
XX
AC ABG20278;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20269.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS84465.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 50637; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2400 AA;
Query Match 15.0%; Score 88; DB 4; Length 2400;
Best Local Similarity 29.3%; Pred. No. 24;
Matches 41; Conservative 19; Mismatches 46; Indels 34; Gaps 9;
QY 4 YVEIERRIENHFLFYKDESIV-----QLQDALSSATTSLTQSN-----NDRGS 50
DB 83 YFEIEKR-----LSHSQERLVNRETRECQSLRLEKLNQKALTEKRNKELEIAQDRNI 136
QY 51 GLLSSGFLR-KQNSNNHSDISNRLTNDLSQELARLKSNLN--NEGMYFTATPSASLEV 107
DB 137 AIQSQFTTRKELEAEKRLD--IRT-NERLSQELEVLTVKRLNEKLESNTTGGELQ- 192
QY 108 IKYD-----LSYLKRRL 119
DB 193 LKLDLQASDVSVKYREKRL 212

```



```
RESULT 8
ABG20276
ID ABG20276 standard; protein; 523 AA.
XX
AC ABG20276;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20267.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
DR N-PSDB; AAS84463.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
Claim 20; SEQ ID NO 50635; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 523 AA;
Query Match 14.6%; Score 85.5; DB 4; Length 523;
Best Local Similarity 29.0%; Pred. No. 5.6;
Matches 36; Conservative 16; Mismatches 43; Indels 29; Gaps 7;
OY 4 YVEIERRIENHFLFYKDSLV-----QLQDALSSATTSLTSALTQSN-----NDRGS 50
DB 54 YFEISKR-----LSHQRLVNETRECOSLEKLNQKALKTERKNELEIAQDRNI 107
OY 51 GLLSSFLR-KQNNHNSKDISNRLTINDLSQLARLKSINLNNEGMFYATPSPASLEVIK 109

RESULT 9
ABG30366
ID ABG30366 standard; protein; 1309 AA.
XX
AC ABG30366;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #30357.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
DR N-PSDB; AAS94553.
XX
New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
Claim 20; SEQ ID NO 60725; 103pp; English.
XX
The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1309 AA;
Query Match 14.3%; Score 84; DB 4; Length 1309;
Best Local Similarity 23.4%; Pred. No. 28;
Matches 39; Conservative 24; Mismatches 50; Indels 54; Gaps 7;
```

QY 1 QKIYVEIERRIENHFLFYKDSLVQ-----LQDAL-----S 32
Db 874 QRYFQIKR---THNFRYFKQLAQOYPEAVPLIYHLQTLLEHPELQLMYANSQKQK 930
QY 33 SATTSLSALTQSNDRGSLSSFL---RKQNNHSHKIDINLRTL-----NDSL 79
Db 931 SLTKMSAKSEANIDRFCELTQSWLPMDKTEKDPNQSSALSSSIRLSGLRINSKADDA 990
QY 80 SOELA-RLKSNLN-----NEGMFYTATPSASLEVIKYDLSYLK 116
Db 991 GOSIANRFTTSNIKGLTQARNANDGISVAQTTEGALSINNLRQR 1037
RESULT 10
ABB61060
ID ABB61060 standard; protein; 1984 AA.
XX ABB61060;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 9972.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR N-PSDB; ABL051163.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 9972; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1984 AA;
Query Match 14.3%; Score 84; DB 4; Length 1984;
Best Local Similarity 30.0%; Pred. No. 50;
Matches 27; Conservative 16; Mismatches 39; Indels 8; Gaps 2;
QY 29 DALSSATTSLALTQSNDRGSLSSFLRKQNNHSHKIDINLRTLNDLSLQELARLKS 88
Db 1142 DANSSNDSSATTSGNCTGNCNQDEAGKEQLQKNNVSKDLRLKLSLEON----LARIQM 1197
QY 89 NLNNEGMFYTATPSASLEVIKYDLSYLKRR 118

Db 1198 MREN---YDAGDBISELLKQMSLFLMQR 1223
RESULT 11
ADX80539
ID ADX80539 standard; protein; 650 AA.
XX ADX80539;
XX 21-APR-2005 (first entry)
XX Plant full length insert polypeptide seqid 49905.
XX plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX Unidentified.
OS US2004034888-A1.
PN 19-FEB-2004.
PD 28-APR-2003; 2003US-00425114.
PF 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
PI WPI; 2004-180133/17.
DR New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX Claim 1; SEQ ID NO 49905; 15pp; English.
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX Sequence 650 AA;

Query Match 14.1%; Score 83; DB 8; Length 650;
 Best Local Similarity 23.4%; Pred. No. 14;
 Matches 40; Conservative 28; Mismatches 49; Indels 54; Gaps 6;
 QY 3 IYVEIERR-----IENHNF-----LFYKDESILVQLQD 29
 Db 325 LVIELNSRVSSLTIIENNFRSCDVMEKYNDLKFSVAKKAALDSKVLREKEEELHQLKD 384
 QY 30 -----ALSATWS---LSALTQNNDRGGLLS-----SFLRKQ-----NSNNHAKD 68
 Db 385 QPEALGASKSQIIVADLSQBRDDLKEALDNESKVNHLKQELQVTOENLAKSRNESAE 444
 QY 69 ISNLTLDNLSLQELARLKSNNLNEGMPYATPSASLEVIKVDLSYLKRL 119
 Db 445 LENLLTLNKLCKELEVLKSLSELTEVNESLQRLNDDAKHEAMELASEL 495

RESULT 12
 ADN23675
 ID ADN23675 standard; protein; 1091 AA.

AC ADN23675;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #6328.
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX
 OS Bacteria.

XX
 XX US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 XX (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX
 XX WPI; 2004-061375/06.
 DR
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 6328; 122pp; English.

XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,

CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 1091 AA;

Query Match 14.1%; Score 83; DB 8; Length 1091;
 Best Local Similarity 22.9%; Pred. No. 28;
 Matches 36; Conservative 22; Mismatches 45; Indels 54; Gaps 6;
 QY 5 VEIERRIENHN-----FLFYKDESILVQLQD-----SSATTSLS 39
 Db 722 IEMERILDNHNKLEKLEKREELKKSHTSHTSLESVLEEQNELAQIQDLREEKEQSNLL 781
 QY 40 ALTO---SNNDRGSGLLSFLRKQNNH---SKDISNLR-----TLNDSLSQELARLKS 88
 Db 782 VLNQKIEKSEKERLEQIRSHTSQNSDTSKTISDLEDKISELLKTNLDLALDVQLKSK 841
 QY 89 NINNEG-----FYTATPSASLEV 107
 Db 842 SLDSKQOLKEADEKNLMLLEVQALQATPSDSAEI 878

RESULT 13
 ABB77431
 ID ABB77431 standard; protein; 761 AA.
 XX
 AC ABB77431;
 XX
 DT 19-JUL-2002 (first entry)
 XX
 DE Human tumour marker protein se2-2.
 KW Human; tumour; cytostatic; cutaneous T cell lymphoma; CTCL; vaccine;
 KW antigen-presenting cell; tumour-specific T cell.
 XX
 OS Homo sapiens.

XX WO200238803-A2.
 XX
 PD 16-MAY-2002.
 XX
 PF 08-NOV-2001; 2001WO-DE004229.
 XX
 XX
 PR 08-NOV-2000; 2000DE-01055285.
 XX
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 XX
 PI Eichmueller S, Schandendorf D, Usener D;
 XX
 XX WPI; 2002-426959/45.
 DR N-PSDB; ABL58958.

XX
 PT Composition containing tumor-associated nucleic acid, useful for
 PT diagnosis and treatment of tumors, especially cutaneous T cell lymphoma,
 PT also derived proteins and antibodies.

PS Claim 5; Fig 8; 84pp; German.

XX The invention relates to a diagnostic composition containing at least one
 CC of 23 nucleotide sequences (I, ABL58901-ABL58950) with altered expression
 CC associated with tumors. (I), including antisense sequences and
 CC ribozymes, also proteins (II, ABB77424-ABB77445) encoded by them and
 CC antibodies specific for (II), are useful for diagnosis, monitoring and

CC treatment of tumours, especially cutaneous T cell lymphoma (CTCL). (II)
 CC and antibodies to (II) are useful for vaccination. (II) can also be used
 CC to prepare pre-loaded antigen-presenting cells or tumour-specific T cells
 XX
 SQ Sequence 761 AA;
 Query Match 14.1%; Score 82.5; DB 5; Length 761;
 Best Local Similarity 24.3%; Pred. No. 19;
 Matches 28; Conservative 32; Mismatches 42; Indels 13; Gaps 4;
 QY 6 EIERIENHNFLFYKDESIVLOQALSSATTSLSALTQSNDRGSGLLSSFLRKQNS-NN 64
 DB 161 ELKTQVEDLN-----ENLLKLEALTKSKNRENSLTDNLND----LNNELQKKQKAYNK 210
 QY 65 HSKDISNLTNLSQELARLKNLNEGMYTATPSASLEVIKYDLSYLKRL 119
 DB 211 ILREKEIDQENDELKQIKRLTSGL--QGKPLTDNKQSLIEELQKVKYKLENQL 263
 RESULT 14
 ADN01130
 ID ADN01130 standard; protein; 1396 AA.
 XX
 AC ADN01130;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Human cell growth, differentiation, and death-associated protein #14.
 XX
 KW human; cell growth; cell differentiation; cell death; CGDD;
 KW cell proliferative disorder; arteriosclerosis; atherosclerosis; bursitis;
 KW cirrhosis; hepatitis; polycythaemia vera; psoriasis; cancer;
 KW developmental disorder; Cushing's syndrome; hypothyroidism;
 KW neurological disorder; epilepsy; stroke; Alzheimer's disease;
 KW Pick's disease; Huntington's disease; Parkinson's disease;
 KW multiple sclerosis; autoimmune disorder; inflammatory disorder; AIDS;
 KW allergy; anaemia; asthma; contact dermatitis; diabetes mellitus;
 KW reproductive disorder; infertility; endometriosis; uterine fibroid.
 XX
 OS Homo sapiens.
 XX
 XN WO2004031364-A2.
 XX
 PD 15-APR-2004.
 XX
 PF 03-OCT-2003; 2003WO-US031441.
 XX
 PR 03-OCT-2002; 2002US-0416205P.
 PR 25-OCT-2002; 2002US-0421521P.
 PR 21-NOV-2002; 2002US-0428376P.
 PR 23-DEC-2002; 2002US-0436258P.
 PR 10-JAN-2003; 2003US-0439292P.
 PR 13-FEB-2003; 2003US-0447578P.
 XX
 PA (INCY-) INCYTE CORP.
 PA (BURR/) BURRILL J D.
 XX
 PI Elliott VS, Swarnakar A, Tang YT, Yue H, Lu DAM, Khare R;
 PI Chawla NK, Richardson TW, Marquis JF, Lal PG, Nguyen DB, Lee SY;
 PI Tran UK, Bhatia UG, Lee S, Blake JJ, Ho A, Zheng W, Gao J, Tran B;
 PI Yang YG, Gietzen KJ, Hafalia AJA;
 XX
 DR WPI; 2004-330172/30.
 DR N-PSDB; ADN01155.
 XX
 XX New isolated polypeptides associated with cell growth, differentiation
 and death, useful for diagnosing, treating or preventing e.g.
 PT atherosclerosis, psoriasis, cancers, Alzheimer's disease, AIDS, anemia,
 PT diabetes mellitus or infertility.
 XX
 PS Claim 1; SEQ ID NO 14; 213pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human

CC proteins that are associated with cell growth, differentiation, and death
 CC (CGDD). The DNA and protein sequences of the invention are useful for
 CC diagnosing, treating or preventing disorders associated with aberrant
 CC expression of CGDD, such as: cell proliferative disorders (e.g.
 CC arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis,
 CC polycythaemia vera, psoriasis and cancers), developmental disorders (e.g.
 CC Cushing's syndrome and hypothyroidism), neurological disorders (e.g.
 CC epilepsy, stroke, Alzheimer's disease, Pick's disease, Huntington's
 CC disease, Parkinson's disease and multiple sclerosis),
 CC autoimmune/inflammatory disorders (e.g. AIDS, allergies, asthma,
 CC contact dermatitis and diabetes mellitus), and reproductive disorders
 CC (e.g. infertility, endometriosis and uterine fibroid). The present amino
 CC acid sequence represents a human CGDD-associated protein of the
 CC invention.
 XX
 SQ Sequence 1396 AA;
 Query Match 14.1%; Score 82.5; DB 8; Length 1396;
 Best Local Similarity 24.3%; Pred. No. 44;
 Matches 28; Conservative 32; Mismatches 42; Indels 13; Gaps 4;
 QY 6 EIERIENHNFLFYKDESIVLOQALSSATTSLSALTQSNDRGSGLLSSFLRKQNS-NN 64
 DB 846 ELKTQVEDLN-----ENLLKLEALTKSKNRENSLTDNLND----LNNELQKKQKAYNK 895
 QY 65 HSKDISNLTNLSQELARLKNLNEGMYTATPSASLEVIKYDLSYLKRL 119
 DB 896 ILREKEIDQENDELKQIKRLTSGL--QGKPLTDNKQSLIEELQKVKYKLENQL 948
 RESULT 15
 ABG20279
 ID ABG20279 standard; protein; 2415 AA.
 XX
 AC ABG20279;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #20270.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS84466.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 50638; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 2415 AA;

Query Match 14.0%; Score 82; DB 4; Length 2415;
Best Local Similarity 28.6%; Pred. No. 1.1e+02;
Matches 40; Conservative 19; Mismatches 47; Indels 34; Gaps 9;

QY 4 YVEIERRIENHNFYKDESIV-----QLQDALSSATTSLTQSN-----NDRGS 50
Db 116 YFEIEKR-----LSHSQRLVNETRECQSLRELEKLNQKALPERKKELEIAQDRNI 169

QY 51 GLLSSFLR-KQSNNHKSDISNLTINDLSQELARLKSNLN--NEGMYTATPSASLEV 107
Db 170 AIQSQFTRTKLEAEKRDLD--IRT-NERLSQELVLTEDVKRLNEKLESNTTKGELQ- 225

QY 108 IKYD-----LSYLKRRL 119
Db 226 LKLDLQASDVSVKYREKRL 245

Search completed: January 24, 2006, 19:44:23
Job time : 29.5934 secs

This Page Blank (uspto)

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 407 AA;

Query Match 100.0%; Score 2062; DB 8; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.8e-155;
 Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCDLAKTKLIDKISSLELYKYKIFFRNYIENVAEDCLKNGLLLESAAHNVSEVELARLKV 60
 DB 1 MCDLAKTKLIDKISSLELYKYKIFFRNYIENVAEDCLKNGLLLESAAHNVSEVELARLKV 60

QY 61 QLNKALLNCIIISYRFGHGYVLVKTDLIDLEQPNVIELPIGFFYLDYEVYRDIGVDFD 120
 DB 61 QLNKALLNCIIISYRFGHGYVLVKTDLIDLEQPNVIELPIGFFYLDYEVYRDIGVDFD 120

QY 121 HITYKVKSNKNSNDVAKVHKSRLLIYENFDYILKRYVPCYTESFLDLYLFEKIYVEI 180
 DB 121 HITYKVKSNKNSNDVAKVHKSRLLIYENFDYILKRYVPCYTESFLDLYLFEKIYVEI 180

QY 181 ERIENHNFLFYKDESIVQLQDALSATTSLTSALTQSNNDRGSGILSSFLRKQNSNNHSHK 240
 DB 181 ERIENHNFLFYKDESIVQLQDALSATTSLTSALTQSNNDRGSGILSSFLRKQNSNNHSHK 240

QY 241 DISLNRNLNDSLSQELARLKSNNLNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300
 DB 241 DISLNRNLNDSLSQELARLKSNNLNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300

QY 301 ADTKPLTRSFNEQAKGNDGKDRSNYYDFLKGVQEQVENACNLKTKYFGLDMKFN 360
 DB 301 ADTKPLTRSFNEQAKGNDGKDRSNYYDFLKGVQEQVENACNLKTKYFGLDMKFN 360

QY 361 LIMLSEEQKVERDIKLIELYSKYNOLIOSSSFDNEELAILKEKLFSP 407
 DB 361 LIMLSEEQKVERDIKLIELYSKYNOLIOSSSFDNEELAILKEKLFSP 407

RESULT 2
 ADU98743
 ID ADU98743 standard; protein; 119 AA.
 AC ADU98743;
 XX
 XX
 XX
 XX 24-FEB-2005 (first entry)
 DE Borrelia burgdorferi antigenic polypeptide seqid 2.
 XX
 XX antibacterial; vaccine; immune stimulation; immunity; antigen;
 KW DNA library.
 XX
 XX Borrelia burgdorferi.
 OS
 XX WO2004103269-A2.
 PN
 XX
 XX 02-DEC-2004.
 PD
 XX 17-OCT-2003; 2003WO-US033056.
 PF
 XX 18-OCT-2002; 2002US-0419401P.
 PR
 XX (MACR-) MACROGENICS INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 XX Sykes KF, Hale KS, Johnston SA;
 PI WPI; 2004-834155/82.
 XX DR N-PSDB; ADU98742.
 XX

PT Immunizing a subject against Borrelia burgdorferi infection comprises
 PT providing to the subject at least one Borrelia antigen or its fragment.
 XX
 XX Claim 27; SEQ ID NO 2; 121pp; English.

CC The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide;
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 119 AA;

Query Match 27.4%; Score 566; DB 8; Length 119;
 Best Local Similarity 95.8%; Pred. No. 4.2e-37;
 Matches 114; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 174 EKIVVEIERRIENHNFLFYKDESIVQLQDALSATTSLTSALTQSNNDRGSGILSSFLRKQ 233
 DB 1 QKIVVEIERRIENHNFLFYKDESIVQLQDALSATTSLTSALTQSNNDRGSGILSSFLRKQ 60

QY 234 NSNNHSHKDISLNRNLNDSLSQELARLKSNNLNEGMYTATPSASLEVIKYDLSYLKEAL 292
 DB 61 NSNNHSHKDISLNRNLNDSLSQELARLKSNNLNEGMYTATPSASLEVIKYDLSYLKRRL 119

RESULT 3
 AAG82132
 ID AAG82132 standard; protein; 583 AA.
 XX
 XX AAG82132;
 XX
 XX
 XX 03-SEP-2001 (first entry)
 DT
 XX
 XX S. epidermidis open reading frame protein sequence SEQ ID NO:1358.
 DE
 XX Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
 KW endocarditis.
 XX
 XX Staphylococcus epidermidis.
 OS
 XX WO200134809-A2.
 PN
 XX 17-MAY-2001.
 PD
 XX 09-NOV-2000; 2000WO-US030782.
 PF
 XX 09-NOV-1999; 99US-0164258P.
 PR
 XX (GLAX) GLAXO GROUP LTD.
 PA
 XX Kimmerly WJ;
 PI
 XX WPI; 2001-316495/33.
 DR N-PSDB; AAH52982.
 XX
 XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.
 XX
 XX Claim 18; Page 386; 2188pp; English.
 XX

XX 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 2070; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX Sequence 1957 AA;
Query Match 7.3%; Score 150.5; DB 8; Length 1957;
Best Local Similarity 20.5%; Pred. No. 0.024;
Matches 96; Conservative 82; Mismatches 183; Indels 107; Gaps 17;
QY 3 DLRKTKLIDKISSLELYKYSIFP-----RNVIENVAEDCLKNGILLESAAHN 49
DB 697 DLRKSEALRFKLEAKNLREVIDNLKGHETLEAQRDLHSLSDAKNTAIISSLETGK 756
QY 50 VSEVELARLKVO-----LKNALLNCIISYR-----FHGI--GYVLVTKT-DTLID 91
DB 757 SSE-DVKRLTANVEITQDSKAMQSFSLVNSYQISNLNLYHELRDDHVMQSQNTLLE 815
QY 92 LRQPV-----NIELPIGEYLDYEVRLDGVDPDHITYKVKSNKNNKNSLDVAKIH 141
DB 816 SESKLTDCENLTQQNMTLIDNVQKLMKHVNO-----ESKVSSELKVNKLSLDLKNLR 870
QY 142 KSRLLIYENFDYIL-----KRYVPCYTESFLDYLFEKLYVEIERRIENHNFLPYKD 194
DB 871 SLSLNAISNDQILTLQABLSEKNYDSLEQESQLNSGL-----KSLEAKQLLHTE 921
QY 195 ESIVOLQDALSATTSLSALTQSNDRGSGILSSFLRKQNSNNHSDISNLNLDLSQ 254
DB 922 NE--ELHRLDKLTGKLTERRSKSSDLGKLTA-----RQETSINLKEENWSQSQ 969
QY 255 ELARLKSNNLNEGMFTTAFPSLASLEVIKYDLSYLVKALALIKAKIGADTKPPLTRSFNBQ 314
DB 970 AITSVKSCLDE-----TLSKSKLEA---DIEHLKNKVSEVEVE-----RNALLASNERL 1016

QY 315 AKGLGNDG-----KGRSNYYDFLKGVQEQVENACNLKLTXYFGLDMKFNLSL 361
DB 1017 MDDLKNGENIASLOTIEKKRAENDDLQSKLSVVSSEYENLLISSQTNSKLEDKTNQL 1076
QY 362 IMLSE-----EQKVERDIKLIYELYSKYNQLIQSSFDNEELAIKKEK 403
DB 1077 KYIEKNVQKLIDEXQNRNVEELTYSKYGLGEENAQIKDELLARKK 1124
RESULT 9
ADN18442
ID ADN18442 standard; protein; 1169 AA.
XX
AC ADN18442;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #1095.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 1095; 122pp; English.
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress

XX	SQ	Sequence 642 AA;	
	Query Match	6.7%; Score 138; DB 2; Length 642;	
	Best Local Similarity	21.5%; Pred. No. 0.051;	
	Matches 104; Conservative	76; Mismatches 162; Indels 142; Gaps 25	
Qy	24	FFRNYIENVAEDCLKNGLIILESAAHNVSEVELARKVOLKNALLNCIIISYRPHGIGYVLV 83	
Db	91	FIQCIRQLCEFLTENG-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143	
Qy	84	KTQDTLIDLEQPNVNIELPIGFEYLDYEV-----VRDLGV--DPDHIIYKVKSNKNKNSL 135	
Db	144	DTK-----FEE-----EVPRIKDLGYPFALSKSSMYTVGAPHTWPHIVAALVW-----LI 189	
Qy	136	DAVKIH--KSLRIIYEN-----FDYILKRYVPCVYTESFLLDIYLFE 174	
Db	190	DCIKIHFMKSSSLFDGQGWGEETEDGIMHNKLFIDYTIK-----CY-ESFMSGADSF 244	
Qy	175	KIVVEIERRIENHNFIFYKDESLVQLDALSSATTSLSALTQSNNDRGSGILSSFLRK-- 232	
Db	245	EMNAELQSKLD--LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLS--LRKLK 300	
Qy	233	-----QSNNHSSKDI--SNURLN-----DSLSEIARLKSNLNNEG 267	
Db	301	ASLQGDVQKYQAVMSNLSSHSAILDQKLNGLNEETARVELECETIKQENTRLQNIIDNQK 360	
Qy	268	MFYTATPSASLEVIKYDLISYLKEALALIKAKIGADTKPLTRSFNEQAK--GLGNDG---- 322	
Db	361	--YSV---ADIERINHNERNELQQT-----NKLTQDLAEQOKLMNEELUYARGKEAIE 411	
Qy	323	-----KG--DRSNYDF-----LKGVQEQVNA 343	
Db	412	LAEVHKLARKLKLIPKGAENSGDYDFEIKFNPEAGANCLVYKRAQVYVYPLKELLNETEE 471	
Qy	344	CNLKLTKYVGLDMKFNSL--IMLSEQKVERDIK--LIELYSKYNQLIOSSSFDNEELAIL 400	
Db	472	INKALNKQMGLEDTLLEQLNAMITTESKRSVRTLKERVQKLDLIIYQOKIKEABEEDKCASE 531	
Qy	401	KEKL 404	
Db	532	LESL 535	
RESULT 14			
ABU03917	ID	ABU03917 standard; protein; 642 AA.	
XX	AC	ABU03917;	
XX	DT	29-JAN-2003 (first entry)	
XX	DE	Human expressed protein tag (BPT) #583.	
XX	KW	Translational profiling; expressed protein tag; EPT; kinase; phosphatase;	
XX	KW	protease; protease inhibitor; transporter; cytoskeletal protein;	
XX	KW	receptor; transcription factor; cancer; MHC;	
XX	KW	major histocompatibility complex; myeloma; colon cancer; gastric cancer;	
XX	OS	adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.	
XX	PN	Homo sapiens.	
XX	FN	W0200278524-A2.	
XX	PD	10-OCT-2002.	
XX	PF	28-MAR-2002; 2002WO-US009671.	
XX	PR	28-MAR-2001; 2001US-0279495P.	
XX	PR	21-MAY-2001; 2001US-0292544P.	
XX	PR	08-AUG-2001; 2001US-0310801P.	
XX	PR	01-OCT-2001; 2001US-0326370P.	
XX	PR	04-DEC-2001; 2001US-0336780P.	

Db	532	LESL	535
Db	91	FIQCIRQLCEFLTENG	-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143
Qy	84	KTKDTLIDLEQPVNIELPIGEYLDY	-----VRDLGV--DPDHIYKVKNNKNSL 135
Db	144	DTK-----FEE-----	EVPRIFKDLGYPFALSKSSMYTVGAPHTWPHIIVAALVW-----LI 189
Qy	136	DAVKIH--KSRLIYEN-----	FDYILKRYVPCYTESFLLDIYLPE 174
Db	190	DCIKIHAMKSSPLFDDGQGWGETEDGIMHINKFLDYTIK	-----CY-ESFMSGADSF 244
Qy	175	KIYVEIERRIENHNFYKDESIVQLQDALSSATTSLSALTQSNDRSGSILSSFLRK	-- 232
Db	245	EMNAELQSKLD--LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES	-LRKLIK 300
Qy	233	-----QNSNNHSDI-SNIRNLN-----	DSLQELARLKSNNNEG 267
Db	301	ASLQGDVQYQVMSNLESHSAILDQKLNGLNEETARVELECETIKQENTLQNIQK	360
Qy	268	MFYTATPSASLEVIKYDLSYLKEALALIKAGIDTKEPLTRSFNEQAK-GLGNDG	--- 322
Db	361	--YSV--ADIERINHERNELQOTI-----	NKLTQDLAEAOQKLWNEELKYARGKEAIEQ 411
Qy	323	-----KG-DRSNYYDF-----	LKGVBQVQVNA 343
Db	412	LAEVHKLARKLKLIPKGAENSGYDFEIKFNP	PEAGANCLVKYRAQVYVPLKELLNETEE 471
Qy	344	CNLKLTKYFGLDKFNLSL-IMLSEBQKVERDIK--	LIELYSKYNQLIOSSPDNELAIL 400
Db	472	INKALNKKMGLEDLTLEQLNAMITESKRSVRTLKEEVQKLDLQYQKIKEABEEDKCASE	531
Qy	401	KEKL	404
Db	532	LESL	535

Search completed: January 24, 2006, 19:44:26
Job time : 93.9538 secs

Db	532	LESL	535
Db	91	FIQCIRQLCEFLTENG	-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143
Qy	84	KTKDTLIDLEQPVNIELPIGEYLDY	-----VRDLGV--DPDHIYKVKNNKNSL 135
Db	144	DTK-----FEE-----	EVPRIFKDLGYPFALSKSSMYTVGAPHTWPHIIVAALVW-----LI 189
Qy	136	DAVKIH--KSRLIYEN-----	FDYILKRYVPCYTESFLLDIYLPE 174
Db	190	DCIKIHAMKSSPLFDDGQGWGETEDGIMHINKFLDYTIK	-----CY-ESFMSGADSF 244
Qy	175	KIYVEIERRIENHNFYKDESIVQLQDALSSATTSLSALTQSNDRSGSILSSFLRK	-- 232
Db	245	EMNAELQSKLD--LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES	-LRKLIK 300
Qy	233	-----QNSNNHSDI-SNIRNLN-----	DSLQELARLKSNNNEG 267
Db	301	ASLQGDVQYQVMSNLESHSAILDQKLNGLNEETARVELECETIKQENTLQNIQK	360
Qy	268	MFYTATPSASLEVIKYDLSYLKEALALIKAGIDTKEPLTRSFNEQAK-GLGNDG	--- 322
Db	361	--YSV--ADIERINHERNELQOTI-----	NKLTQDLAEAOQKLWNEELKYARGKEAIEQ 411
Qy	323	-----KG-DRSNYYDF-----	LKGVBQVQVNA 343
Db	412	LAEVHKLARKLKLIPKGAENSGYDFEIKFNP	PEAGANCLVKYRAQVYVPLKELLNETEE 471
Qy	344	CNLKLTKYFGLDKFNLSL-IMLSEBQKVERDIK--	LIELYSKYNQLIOSSPDNELAIL 400
Db	472	INKALNKKMGLEDLTLEQLNAMITESKRSVRTLKEEVQKLDLQYQKIKEABEEDKCASE	531
Qy	401	KEKL	404
Db	532	LESL	535

Search completed: January 24, 2006, 19:44:26
Job time : 93.9538 secs

Query Match 6.7%; Score 138; DB 6; Length 642;
Best Local Similarity 21.5%; Pred. No. 0.051;
Matches 104; Conservative 76; Mismatches 162; Indels 142; Gaps 25;

QY 24 PFRNYIENVAEDCLNGLILESAAHNVSEVELARLVQLKALLNCIISYRPHGIGYLV 83

Sequence 642 AA;

Query Match 6.7%; Score 138; DB 6; Length 642;
Best Local Similarity 21.5%; Pred. No. 0.051;
Matches 104; Conservative 76; Mismatches 162; Indels 142; Gaps 25;

QY 24 PFRNYIENVAEDCLNGLILESAAHNVSEVELARLVQLKALLNCIISYRPHGIGYLV 83

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006. Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 71.5326 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-4
Perfect score: 2062
Sequence: 1 MCDLRKTKLIDKISSLELYK.....QSSSFDNEELAILKEKLFSP 407

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2062	100.0	407	5	US-10-688-058-4
2	566	27.4	119	5	US-10-688-058-2
3	157.5	7.6	1009	4	US-10-282-122A-71141
4	157.5	7.6	1010	4	US-10-724-972A-4177
5	150.5	7.3	1957	4	US-10-369-493-2070
6	150.5	7.3	1957	5	US-10-732-923-3328
7	143.5	7.0	1169	4	US-10-369-493-1095
8	141.5	6.9	2910	5	US-10-732-923-3342
9	141	6.8	819	4	US-10-282-122A-47000
10	141	6.8	1527	5	US-10-732-923-3354
11	140.5	6.8	864	5	US-10-732-923-3312
12	140.5	6.8	921	5	US-10-732-923-3305
13	140	6.8	703	4	US-10-282-122A-63529
14	138	6.7	642	4	US-10-648-593-172
15	138	6.7	642	4	US-10-755-889-608
16	138	6.7	642	5	US-10-733-878-441
17	138	6.7	642	5	US-10-473-127-583
18	138	6.7	642	5	US-10-473-127-584
19	138	6.7	642	5	US-10-473-127-585
20	138	6.7	642	5	US-10-473-127-587
21	138	6.7	642	5	US-10-473-127-588
22	138	6.7	642	5	US-10-287-436A-557
23	138	6.7	664	5	US-10-500-530-50
24	135.5	6.6	696	4	US-10-408-765A-1067
25	134	6.5	1939	5	US-10-732-923-3340
26	133.5	6.5	2503	5	US-10-828-985A-11
27	133.5	6.5	2543	5	US-10-828-985A-9

ALIGNMENTS

RESULT 1

US-10-688-058-4
; Sequence 4, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCR0:00305
; CURRENT APPLICATION NUMBER: US/10/688.058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-10-688-058-4

Query Match	100.0%	Score 2062;	DB 5;	Length 407;
Best Local Similarity	100.0%	Pred. No. 5.8e-149;		
Matches 407;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLKNGLILESAAHNVSEVELARLKV	60	
Db	1	MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLKNGLILESAAHNVSEVELARLKV	60	
Qy	61	QLKVALNCIIISYRPHGIGYLVKTKOTLIDLEQPNIELPFGFYLDYEVYRDIGVDPD	120	
Db	61	QLKVALNCIIISYRPHGIGYLVKTKOTLIDLEQPNIELPFGFYLDYEVYRDIGVDPD	120	
Qy	121	HITTKVKNNNKNSLDVAKIHKSRLLIYENFDYILKRVPCVCTESFLDIDYLFKEIYVEI	180	
Db	121	HITTKVKNNNKNSLDVAKIHKSRLLIYENFDYILKRVPCVCTESFLDIDYLFKEIYVEI	180	
Qy	181	ERRIENHNFLYKDESLVQLQDALSSATTSLSALTQSNDRGSGILSSFLRKQSNHNSK	240	
Db	181	ERRIENHNFLYKDESLVQLQDALSSATTSLSALTQSNDRGSGILSSFLRKQSNHNSK	240	
Qy	241	DISLNRLNLSLSQELARLKSNNLNNEGMYFATPSASLEVIKYDLSYLKEALALIKAGI	300	
Db	241	DISLNRLNLSLSQELARLKSNNLNNEGMYFATPSASLEVIKYDLSYLKEALALIKAGI	300	
Qy	301	ADTKEPLTRSFNEQAKGLGNDGKDRSNYDPLKGVQSOVENACNLKLTKYFGLDMKFN	360	
Db	301	ADTKEPLTRSFNEQAKGLGNDGKDRSNYDPLKGVQSOVENACNLKLTKYFGLDMKFN	360	

28	133.5	6.5	2568	5	US-10-828-985A-7	Sequence 7, Appli
29	133.5	6.5	2633	5	US-10-450-763-36864	Sequence 36864, A
30	133.5	6.5	2663	5	US-10-723-860-749	Sequence 749, App
31	131.5	6.4	878	4	US-10-724-972A-4767	Sequence 4767, Ap
32	131.5	6.4	1184	4	US-10-282-122A-53254	Sequence 53254, A
33	129.5	6.3	365	4	US-10-125-692-16	Sequence 16, Appl
34	129.5	6.3	365	5	US-10-991-347-16	Sequence 16, Appl
35	129.5	6.3	575	4	US-10-601-036-8	Sequence 8, Appli
36	129.5	6.3	1182	4	US-10-282-122A-53445	Sequence 53445, A
37	128.5	6.2	996	3	US-09-815-242-5251	Sequence 5251, Ap
38	128.5	6.2	1009	3	US-09-815-242-12141	Sequence 12141, A
39	128.5	6.2	1009	4	US-10-282-122A-43832	Sequence 43832, A
40	126	6.1	980	5	US-10-732-923-3343	Sequence 3343, Ap
41	126	6.1	2133	5	US-10-732-923-15030	Sequence 15030, A
42	125	6.1	436	4	US-10-120-801-88	Sequence 68, Appl
43	125	6.1	436	4	US-10-275-595A-16	Sequence 16, Appl
44	125	6.1	1478	5	US-10-732-923-3353	Sequence 3353, Ap
45	125	6.1	1875	4	US-10-369-493-22285	Sequence 22285, A

		LIMLSEEQVERDIKLI	LYSKYNQLIQSSFDNEELAILKEKLPSF	407
Qy	361			
		LIMLSEEQVERDIKLI	LYSKYNQLIQSSFDNEELAILKEKLPSF	407
Dd	361			
		LIMLSEEQVERDIKLI	LYSKYNQLIQSSFDNEELAILKEKLPSF	407

RESULT 2

```

US-10-688-058-2
; Sequence 2, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-2

```

RESULT 3

```

US-10-282-122A-71141
; Sequence 71141, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of En
; FILE REFERENCE: EUIPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,1
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

RESULT 4

```

US-10-724-972A-4177
; Publication 4177, Application US/10724972A
; Sequence No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGN
; FILE REFERENCE: PAT#03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 4177
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: S.epidermidis
US-10-724-972A-4177

```

```

Query Match          7.6%; Score 157.5; DB 4; Length 1010;
Best Local Similarity 25.6%; Pred. No. 0.0044;
Matches 80; Conservative 55; Mismatches 98; Indels 79; Gaps 16;

QY 131 KNSLDVAKHKSRLLIYENDYILKRYVPCYTSFLDIYLFKIVVEIERR--LEN- 186
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 162 QNSKDKQSILRT-LFNSERFDEI-----RHLIVENVKQEK--VQIENRYTQIENL 209
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 187 -HNFLFYKDESIVQIQDALSGATTSLSALTOQNDRSGSILSSFLRKQNSNHSKDISNL 245
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 210 WNDIDTFNNDELALYKELESSQTDKWEKFFQFNDYGCKILKSP--EAKNKITKELDDL 267
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 246 RN--LNSLSQELARLKNLSNNEGMYTATPSASLEVIKYD-----LSYIKEALAL 294
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 268 NHKYKVNVELSENTKKLA-----EKIKPDDLKKEQNYIDKLIKQELKLM 310
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 295 I---KAKIGADTKPELITPSFNEAQKLGNDGKDSNYYDFLKGVOQOVENACN----- 345
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 311 IQESKVLITYFTRLOSQKDKDELVSLEHQSKINATNYHNEIKGFQKQLEHLSTRENEIT 370
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 346 ----LKLTKYFGLDMKFNSL--IMLSEQK--VERDIKLIYELSKYNOLI-----QS 389
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 371 QFNQYLERKNQVF-----FNQLDKIISYQCKPVIIEEIK--RLYSEYNDLITKKEELTKE 423
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 390 SSFDNEELAILK 401
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 424 MNKNKDFALIE 435
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

RESULT 5
US-10-369-493-2070
; Sequence 2070, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2070
; LENGTH: 1957
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2070

Query Match          7.3%; Score 150.5; DB 4; Length 1957;
Best Local Similarity 20.5%; Pred. No. 0.037;
Matches 96; Conservative 82; Mismatches 183; Indels 107; Gaps 17;

QY 3 DLKRTKLDIKISLELYKYSIFF-----RNYIENVAEDCLKNGLILSAAHN 49
    |||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 697 DLKRSSEALRFSKLEAKNIREVIDNLKGHETLEQRNDLSLSDAKNTVAIISSELTK 756
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 50 VSEVELARLKVQ-----LKNALLNCIISYR-----FHGI--GYVLVKTK-DTLID 91
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 757 SSB-DVKRLTANVETLTQDSKAMQSFSLVNSYQISINLYHELDRDDHVNMQSQNTLLE 815
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 92 LEQPV-----NIEPIGEYLDYEVYRDLGVDFDHIYTYKYSKNNKNSLDVAKIH 141
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 816 SESKLTDCENLTQNNWTLIDNVQKLWHKVNQ-----ESKVSSELKVEYNGKLSLDLKNLR 870
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
QY 142 KSRLLIYENPDYIL-----KRYVPCYTESFLDIYLFKIVVEIERIENHNPLFYKD 194
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
DB 871 SSNLVAISNDQILTLQAEIASKNYDSLEQSAQLNSGL-----KSLAEKQLHTE 921
    :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

```

```

RESULT 7
US-10-369-493-1095
; Sequence 1095, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1095
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-10-369-493-1095

Query Match          7.0%; Score 143.5; DB 4; Length 1169;
Best Local Similarity 22.2%; Pred. No. 0.063;
Matches 113; Conservative 87; Mismatches 167; Indels 141; Gaps 29;

QY 4 LRKTLIDKISLELYKYISFRNVIENVAEDCLKNGL-ILESAAHNVSEVE--LARKV 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 159 TERRKIIDEISGIAEFDEK-----KKKAEELKARELIMDIRISEVENNLKLLK 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 QLKALLNCIISYRFGIGYLVKTKDILID--LEQPVNIELPI-GFEYLDVEY---VRD 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 212 EKEDAEKIKUNEELKAAYALILKVSYLNVLE---NIQNDIKNEELKNEFLSKVRE 268
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 LGVDPDHITYKYVKNKNNKSL-----DAVKIHSRLIYENFDYILKRYVPCYTESPLL 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 269 IDVEIENL--KLNNIINELNEKNEVEVLELKS-----IKELEVEIENDKKVL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 D--IVLFKIVYVEIRRR-----IENHFLFYKDESIVQ-----LQ 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 317 DSSINELKKEVEIENKKEIKETQKTIENRDSIIEKEQIKETEKKIKNLNVEKERLK 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 202 DALSSATTSLSALTQSNDRSGILSSFLRKONS-----NNHSKOISLNLND----- 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 377 EAIASESIIKHLKESEME-----IADETAKNQNELYRLKKELDLDLNLNRKNFIEKNN 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 251 ----SLSQELARLK-----SNLNNMGFYTATPSASLEVIKYDL-----S 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 EMIKKLKELEIETVEDVTKPLYLEENLVNVE-IEFSKGIKELEKKEKQKLDLHAE 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 YLKEALAIKA-----KITGADTKPELTSFNEQAKGL-----GNQDK-----G 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 492 YVKN-ARIKALKEMEELSMD--RAIRELLNANLPGIIDIIVGNLCKTKIEYKTAIEVAAG 548
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 DRSNYVDYDLKGVOEVENACNLKLTG-----YFGILD-MKFNSLIMLSEQKVERDIKLE 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 549 NRLNHI-VVKRMDDAVRAIKYKLERKLGKRGATPLDRIEGREAYVIDEDGVIGRAIDLVE 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 LYSKYNQLIQSSSFON-----BELALKE 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 608 PDEKRVYFE-YVFGNTVVVENIDIKE 634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-732-923-3342
; Sequence 3342, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C

```

```

; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3342
; LENGTH: 2910
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-3342

Query Match          6.9%; Score 141.5; DB 5; Length 2910;
Best Local Similarity 19.2%; Pred. No. 0.3;
Matches 81; Conservative 83; Mismatches 162; Indels 95; Gaps 18;

QY 11 DKISLELYKYISFRNVIENVAEDCLKNGLILSAAHNVSEVELARUKVLKNALLNCI 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 299 NNIDNKELY----YFNESINEIHENCTDSKILILKICNIIFE-----KKTKN----- 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 ISYRFGIGYLVKTKDILIDLEQPVNIELPIGFEYLDVEYVRDLGVDFDHITYKYVKN 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 343 ----IH-----ILKQLDLLFLQK-----SYENRNIFYEKKSLYGMENISSKNINN 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 KNNSLDAVKIHSRLIYENFDYILKRYVPCYTESFLDIYLFKVIYVEIERRIENHNP 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 386 DNNNNN-----YHNFSPGKENVFLDNNNTSINIVREDE---SKEKSI SNYF- 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 FYKDESLVOLQDALSSATTSLSALTQSNND-----RSGGILSSFLRKQNNHSHKDIS 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 433 ---DKTLFLFVLLKVSHTISGLYNTFMDITLNTNSDALNNHINSNNNNNNNNNN 489
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 NLND-SLSQELARLKSLNNEGMFYTATPSASLEVIKYDLSYLK-----EALALIKAKI 299
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 490 NNDYFISYEURLKIESIN-----YQRAVNAALYLFHILHMLLETTSQNNNNMLIENEI 545
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 GADTKPELTSFNE-----QAKGLGNDGKGRSNNYDFDLKGVOEVE----- 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 546 --STIRNFIKINKKKVDSNINILNIYQYKNVPTQNNNTYKSNKQLNIQNNENK 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 NACNLKTYFGLDMKNSLIMLSEQKVERDIK-----LIELYKYNQLIQSSSFONE 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 604 NLINKELVRH---NFEIRKIIICKYNNKIIKWMKKIHNLYIMHYVYSYKN--DQNGYINE 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 E 396
DB 659 E 659

RESULT 9
US-10-282-122A-47000
; Sequence 47000, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47000
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-47000

Query Match
Best Local Similarity 6.8%; Score 141; DB 4; Length 819;
Matches 117; Conservative 93; Mismatches 152; Indels 164; Gaps 30;

QY 4 LRK--TKLIDKISSLE--LYKSIIFRNYIENVAEDCKLQNGLILSAHNVS---EVELA 56
D 216 LRKQDLSQKLEALEKITLKLINFDLVN---LKSELNLDGLSKFLSSGPKLE 271
QY 57 RLKVLQKLNALLCIIISYRPHGIGYVLTKOTLIDLEQPVNIPIGFEYLDYEVRLIG 116
D 272 FYSFEKNNVNN-----IQLEKDLMLKSKILGLBIKIQLEQKXLTAKN 317
QY 117 VDFPHITYKVNKNKNSLDVAIKHSRLIYENFDYI--LKYVPCYTESF--LLDIVLF 173
D 318 LEDKFKVSKSESEKKNINNNLQNLNLLKKNDFPSLDEINRYTKSPFELVDLIL- 376
QY 174 EKIYVEIERIRIHNHFLFYKD---BSLVQLODALS-----SATTSLSALT 215
D 377 -----SVLRSAKSEEFVLKHEILSLKLFESLSIKYKEIRVNLKIYISKDENLALL 431
QY 216 QSNDRGSGI-----LSSFLRKONS--NNHSDKISNLR-----NLNLSQ---E 255
D 432 R---DKIEPIFQNVNLLKFLKNAKSNLAKETNIERLIEBKTQLQNDVLGELEYE 488
QY 256 LARLKSNL-----NNEGMFYTATPSASLEVIKYDLSYK--EALALIKAGADTKEP 306
D 489 LSKFK--NLDEIKLIDGNLFLFESKN---SLDEELKDL--YLKLENLLEKSDIQLNLSN 543
QY 307 L-----TSFNEQ---AKGLNDGKDRSNYYDFLK-----GVQOVEN--- 342
D 544 IGASKFSSEKFKPNFLNDPK--KTNYEYIKQTEYEFELNLSNVSKSEIENFN 601
QY 343 -----ACNLKLTKEFLDMKNSL-----IMLSE--- 367
D 602 VSNEMSDKFELEBDLATRELLRKEIDAIKLGDVVFNFNIDKREDETKERFEKVSQVEDLK 661
QY 368 -----QKVERDITKILYSKNQLIQSSSFD--NEELAILKEKLF 405
D 662 LSKYSLOKLOKKIK--NEIYKKPRE-----SFDEINKNFSFFPKIIP 701

```

```

RESULT 10
US-10-732-923-3354
; Sequence 3354, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

```

```

; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3354
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1527)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-3354

Query Match
Best Local Similarity 6.8%; Score 141; DB 5; Length 1527;
Matches 97; Conservative 91; Mismatches 178; Indels 108; Gaps 22;

QY 6 KTLIDK---ISSLEL--YKYSI-----FFRNYIENVAE---DCLKNG 40
D 812 KEKLEEKYETQIKETEMKYKIQKEIEKTKQNAEQNFNSKFEKYKENLEKNKNDFINNL 871
QY 41 LI-----LESAHNVS---VELARLVQLKNAALLCIIISYRPHGI-----GY 80
D 872 IIEKNNEIESFPNDIEQKFKEMEKFEENEKLLQNNFNENMKNFHIEBQNKHIENIKKEY 931
QY 81 VLVT-----KOTLIDLEQPVNIPIGFEYLDYEVRLDGVDPDHIYKYKSN 129
D 932 ELIKNNEIEVLKEEMKKNKIQEIE---NVELKLADE--KNKHIDDMKKELENI--YNVEIN 985
QY 130 NKNNSLDAVKHKSRLLIYENFDYILKRYVPCYTESFLDIVLFYKEIYVETERIENHNF 189
D 986 NLKEKINAEQTLAEMLTQRYEYEQKLEIQRNEYEQOQLEIRNEMENKLIQEKESYDL 1045
QY 190 LFYDESIQVQDALSSATTSLSALTQSN-----NDRSGIILSSFLRKQNSNNHS 239
D 1046 GSKSESIQILTDKIEKLKWEIIEINKEKIDLEHNISILNDRERINILNKNLENDLKN 1105
QY 240 KDIN--LRNLD---SLSQELARLKNLNNEGMYTATPSASLEVIKYDLSYKALAL 294
D 1106 EDIYKKNILNDEKLKLEKEIEHIKENGAKG-----SQIREQPADLLQEEIDR 1155
QY 295 IKAKIGADTKEPIT---RSFNEQAKGLNDGKDRSNYYDFLKGV---QEOVENACNLKL 348
D 1156 IK-----KESKEKVIYIQYNE-----INEEYEEKKEYNLDLEKANVSKOLTEKCEENI 1207
QY 349 TKYFGLDMKFNLSLI--MLSEBOKVERDIKLIELYSKNQLIQSSSFDNEELAILK 401
D 1208 QKI-----NEYEDMITKMLENQTVELVTKIQELNEDF--LKKEAFDNEKNDLLK 1255

```

```

RESULT 11
US-10-732-923-3312
; Sequence 3312, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3312
; LENGTH: 864
; TYPE: PRT
; ORGANISM: Sulfolobus solfataricus
US-10-732-923-3312

Query Match
Best Local Similarity 6.8%; Score 140.5; DB 5; Length 864;

```



```
Query Match      6.8%; Score 140; DB 4; Length 703;
Best Local Similarity 21.3%; Pred. No. 0.059;
Matches 105; Conservative 73; Mismatches 140; Indels 176; Gaps 26;

Qy 4 LRKTKLIDKISSLELYKYISIFRNYIENVAEDCLKNGILLESAAHNVSEVELARLVQV-- 61
Db 94 LKRFIIDD-----DISYLFKEPLADSKLDINLAKQIIDFNKTFADFINKLDQDER 146
Qy 62 -----LKNALLNCIISYR-----FHGIGYVLVKT----- 85
Db 147 LISLCEHSLLNKDBEYSTLKTQLINAFISYEKNKILANNKLDPHDL---LTKCNLLSNDN 203
Qy 86 -----KDTLIDLEQPNIELPIGFEYLDYEVVDLGVDPDHIYKYVKNK----- 131
Db 204 DLLNOMSEQFOHILVDFQDTN-----QIQVELIKWL-VTRKNKFLVGDNNQMIYR 254
Qy 132 -----NNSLDAVK-----IHKSLRIIYENFDYILKRYVPCYTESFLDIYLPK----- 175
Db 255 WRGAVNGIITALKHDFNVPKSN-BFFINQNYRCQNIILAVANQIILLIMAYEKQVKTEKN 313
Qy 176 -----IYVEIERRIENH-NFLFYKDESIVQDALSSATTSLTSALTQSNNDRG 222
Db 314 LLFSTLNGDKKPVYFOAE-SVENQANWIFNKIKALNQ-----TEKINPKD 357
Qy 223 SGILSSFLRKQNSNHSDIENLRLNDSLSQ-ELARLKSNNLNEGMFYTATPSASLEVI 281
Db 358 MAIL-----FRK-----NRDITMTWELIEADGTIPLPKQKSYFN-----QLV 394
Qy 282 KYDLSYLKEALALIKAKIGADTKPLTRSFNEQAQKGLNDGK-----GDRSNYYDFLKG 336
Db 395 K-----LQRLVIAISTRNLDIKRAL-----QALKIWSNDLKELWKQSKDTNLPDFL 444
Qy 337 Q-EQVENACNLKTYFGLDKMFNSLIMLSEQKVERDIKLIYLS--KYNQLIQSSSFD 393
Db 445 ELNQKHSSKLKATGY-----FNLLIKLAEQDQI--NLLFTPLFKLVKVDQTIEN--- 492
Qy 394 NEELAILKEKLFSP 407
Db 493 -----LLWKKLTEF 501

RESULT 14
US-10-648-593-172
; Sequence 172, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 172
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-172

Query Match      6.7%; Score 138; DB 4; Length 642;
Best Local Similarity 21.5%; Pred. No. 0.075;
Matches 104; Conservative 76; Mismatches 162; Indels 142; Gaps 25;

Qy 24 FFRNYIENVAEDCLKNGILLESAAHNVSEVELARLVQKNALLNCIISYRPHGIGYVLV 83
Db 91 FIOQCIRQLCEFLTENG-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143
Qy 84 KTKDTLIDLEQPNIELPIGFEYLDYEV-----VRDLGV--DFDHIYKYVKNKNNKNSL 135
Db 144 DTK-----FEB-----EVPRIKDLGYPPALSSKSSMYTVGAPHTWPHIVAALVW-----LI 189
Qy 136 DAVKIH---KRLRIIYEN-----FDYILKRYVPCYTESFSLDIYLF 174
Db 190 DCIKIHTAMKESSPLFDDGQPGWBETEDGIMHNKLFIDYTIK---CY-ESFMSGADSF 244
Qy 175 KIYVEIERRIENHNFYKDESIVQDALSSATTSLTSALTQSNNDRGSGILSSFLRK-- 232
Db 245 EMNAELQSKLDK---LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES--LRK 300

Query Match      6.7%; Score 138; DB 4; Length 642;
Best Local Similarity 21.5%; Pred. No. 0.075;
Matches 104; Conservative 76; Mismatches 162; Indels 142; Gaps 25;

Qy 24 FFRNYIENVAEDCLKNGILLESAAHNVSEVELARLVQKNALLNCIISYRPHGIGYVLV 83
Db 91 FIOQCIRQLCEFLTENG-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143
Qy 84 KTKDTLIDLEQPNIELPIGFEYLDYEV-----VRDLGV--DFDHIYKYVKNKNNKNSL 135
Db 144 DTK-----FEB-----EVPRIKDLGYPPALSSKSSMYTVGAPHTWPHIVAALVW-----LI 189
Qy 136 DAVKIH---KRLRIIYEN-----FDYILKRYVPCYTESFSLDIYLF 174
Db 190 DCIKIHTAMKESSPLFDDGQPGWBETEDGIMHNKLFIDYTIK---CY-ESFMSGADSF 244
Qy 175 KIYVEIERRIENHNFYKDESIVQDALSSATTSLTSALTQSNNDRGSGILSSFLRK-- 232
Db 245 EMNAELQSKLDK---LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES--LRK 300
```

```
Db 144 DTK-----FEB-----EVPRIKDLGYPPALSSKSSMYTVGAPHTWPHIVAALVW-----LI 189
Qy 136 DAVKIH---KRLRIIYEN-----FDYILKRYVPCYTESFSLDIYLF 174
Db 190 DCIKIHTAMKESSPLFDDGQPGWBETEDGIMHNKLFIDYTIK---CY-ESFMSGADSF 244
Qy 175 KIYVEIERRIENHNFYKDESIVQDALSSATTSLTSALTQSNNDRGSGILSSFLRK-- 232
Db 245 EMNAELQSKLDK---LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES--LRK 300
Qy 233 -----QSNNNHSKOI-SMLRLNL-----DSLQSLQELARLKSNNLNEG 267
Db 301 ASLOQDVQKYQAYMSNLSESHAILDQKLNGLNBEIARVELECEITKQENTRLQNIIDNQK 360
Qy 268 MYTATPSASLEVIKYDLSYLKEALALIKAKIGADTKPLTRSFNEQAK-GLGNDG----- 322
Db 361 --YSV--ADIERINHNERNELQQTII---NKLTKDLAEQOKLWNEELKYARGKEAIBTQ 411
Qy 323 -----KG-DRSNYYDF-----LKGVOQOVENA 343
Db 412 LAEYHKLARKLKLIPKGAENSGYDFEIKFNPAGANCLVKYRAQVYVPLKELNETSEE 471
Qy 344 CNLKLTKYFGLDKMFNSL-IMLSEQKVERDIK--LIELYSKYNQLIQSSSFDNBEILAIL 400
Db 472 INKALNKKMGLEDLEQLNNAITKESKRSVRTLKEEVQKLDLYQOKIKEABEEDKCASE 531
Qy 401 KEKL 404
Db 532 LESL 535

RESULT 15
US-10-755-889-608
; Sequence 608, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; TITLE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 608
; LENGTH: 642
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-608

Query Match      6.7%; Score 138; DB 4; Length 642;
Best Local Similarity 21.5%; Pred. No. 0.075;
Matches 104; Conservative 76; Mismatches 162; Indels 142; Gaps 25;

Qy 24 FFRNYIENVAEDCLKNGILLESAAHNVSEVELARLVQKNALLNCIISYRPHGIGYVLV 83
Db 91 FIOQCIRQLCEFLTENG-----AHNVSMKSLQAPSV--KDFLKIFTFLYGLFCPSYELP 143
Qy 84 KTKDTLIDLEQPNIELPIGFEYLDYEV-----VRDLGV--DFDHIYKYVKNKNNKNSL 135
Db 144 DTK-----FEB-----EVPRIKDLGYPPALSSKSSMYTVGAPHTWPHIVAALVW-----LI 189
Qy 136 DAVKIH---KRLRIIYEN-----FDYILKRYVPCYTESFSLDIYLF 174
Db 190 DCIKIHTAMKESSPLFDDGQPGWBETEDGIMHNKLFIDYTIK---CY-ESFMSGADSF 244
Qy 175 KIYVEIERRIENHNFYKDESIVQDALSSATTSLTSALTQSNNDRGSGILSSFLRK-- 232
Db 245 EMNAELQSKLDK---LFNVDAFKLESLEAKNRALNEQIARLEQEREKEPNRLES--LRK 300
```

```

QY 233 -----QNSNNHSDI-SNLRNLN-----DSLQELARLKSNNNEG 267
Db 301 ASLQGDVQYQAYMNSHSAIILQKLNGLNEETARVELECEETIKQENTRLQNIIDNQX 360
QY 268 MPYTATPSASLEVIKYDLSYLKBALALIKAKIGADTKPLTRSFNEQAK-GLGNDG----- 322
Db 361 --YSV---ADIERINHRELQOTI---NKLTKDLAEAQQKLWNEELKYARGKEAIEAQ 411
QY 323 -----KG-DRSNYDF-----LKGVOEQVENA 343
Db 412 LAEYHKLARKKLIIPKGAENSGYDFEIKFNPAGANCLVKYRAQVYVPLKELLNETEE 471
QY 344 CNLKLTKYFGLDMKFNSL-IMLSERQKVERDIK--LIELYSKYNOLIOSSFDNEELAIL 400
Db 472 INKALNKKMGLEDITLEQLNAMITESKRSVRILKEEVQKLDLDLYQQKIKEAEEDEKCASE 531
QY 401 KEKL 404
Db 532 LESL 535

```

Search completed: January 24, 2006, 20:54:28
Job time : 73.5326 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 6.51393 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-4
Perfect score: 2062
Sequence: 1 MCDLRKTKLIDKISSELYK.....QSSSFDNELAILKELKLPFSF 407

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB pep.*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB pep.*
6: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB pep.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	157.5	7.6	583	6	US-10-793-626-1358 Sequence 1358, Ap
2	135.5	6.6	2665	7	US-11-124-368A-214 Sequence 214, Ap
3	135.5	6.6	2668	7	US-11-124-368A-215 Sequence 215, Ap
4	126	6.1	5024	6	US-10-793-626-2964 Sequence 2964, Ap
5	114	5.5	710	6	US-10-467-657-4292 Sequence 4292, Ap
6	113.5	5.5	776	7	US-11-033-039-637 Sequence 637, Ap
7	113.5	5.5	989	6	US-10-821-234-975 Sequence 975, Ap
8	112	5.4	1167	7	US-11-052-554A-121 Sequence 121, Ap
9	111	5.4	976	7	US-11-155-288-20 Sequence 20, Ap
10	107.5	5.2	888	7	US-11-077-550-112 Sequence 112, Ap
11	106.5	5.2	1078	7	US-11-165-221-43 Sequence 43, Ap
12	106.5	5.2	1078	7	US-11-165-226-53 Sequence 53, Ap
13	106	5.1	2710	7	US-11-051-453-41 Sequence 41, Ap
14	105.5	5.1	1732	6	US-10-055-877-147 Sequence 147, Ap
15	104	5.0	698	6	US-10-793-626-2388 Sequence 2388, Ap
16	102.5	5.0	864	7	US-11-077-550-102 Sequence 102, Ap
17	102.5	5.0	865	7	US-11-077-550-100 Sequence 100, Ap
18	102.5	5.0	866	7	US-11-077-550-104 Sequence 104, Ap
19	102.5	5.0	867	7	US-11-077-550-80 Sequence 80, Ap
20	102.5	5.0	876	7	US-11-077-550-82 Sequence 82, Ap
21	102.5	5.0	1420	7	US-11-077-550-110 Sequence 110, Ap
22	102.5	5.0	1614	6	US-10-821-234-903 Sequence 903, Ap
23	101.5	4.9	605	6	US-10-689-742-140 Sequence 140, Ap
24	101	4.9	1189	7	US-11-074-176-134 Sequence 134, Ap
25	101	4.9	1735	7	US-11-040-472-13 Sequence 13, Ap

26	100.5	4.9	1145	6	US-10-793-626-1432 Sequence 1432, Ap
27	100	4.8	465	6	US-10-878-556A-116 Sequence 116, Ap
28	100	4.8	495	6	US-10-821-234-1154 Sequence 1154, Ap
29	99	4.8	1095	6	US-10-793-626-3154 Sequence 3154, Ap
30	98.5	4.8	1216	6	US-10-873-528-12 Sequence 12, Appl
31	97.5	4.7	404	6	US-10-793-626-1130 Sequence 1130, Ap
32	97.5	4.7	860	7	US-11-077-550-175 Sequence 175, Ap
33	97.5	4.7	862	7	US-11-077-550-94 Sequence 94, Appl
34	97.5	4.7	866	7	US-11-077-550-88 Sequence 88, Appl
35	97.5	4.7	867	7	US-11-077-550-96 Sequence 96, Appl
36	97.5	4.7	867	7	US-11-077-550-98 Sequence 98, Appl
37	97.5	4.7	870	7	US-11-077-550-92 Sequence 92, Appl
38	97.5	4.7	871	7	US-11-077-550-84 Sequence 84, Appl
39	97.5	4.7	871	7	US-11-077-550-86 Sequence 86, Appl
40	97.5	4.7	871	7	US-11-077-550-90 Sequence 90, Appl
41	97.5	4.7	1070	7	US/11/062 Sequence 7, Appli
42	97.5	4.7	1095	7	US/11/062 Sequence 4, Appli
43	97.5	4.7	1169	7	US-11-077-550-20 Sequence 20, Appl
44	97	4.7	299	7	US-11-156-084-288 Sequence 288, Ap
45	97	4.7	568	6	US-10-793-626-2482 Sequence 2482, Ap

ALIGNMENTS

RESULT 1
US-10-793-626-1358
; Sequence 1358, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIORITY FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1358
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1358

Query Match	7.6%	Score 157.5;	DB 6;	Length 583;
Best Local Similarity	25.6%	Pred. No. 0.00046;		
Matches	80;	Conservative 55;	Mismatches 98;	Indels 79;
Gaps	16;			
QY	131	KNNSLDVAKIKHSRLIIYENFDYILKRYVPCYTESFLDLIVLFEKIVYIEERR----	IEN-186	
Db	154	QNSKDKQSILRT-LFNSERFDEI-----RHLLVENVKQEK--VQIENRYQIENL	201	
QY	187	-HNFIFYKDESIVQIQDALSSATTSLSALTQSNDRGSGILSSFLRKQNSNHSKDINL	245	
Db	202	WNIDITFNDELALYKELESSQTDKMBKFPQFNDYGCKILKSF--EEAKNKITTELDD	259	
QY	246	RN--LNLSLSQELARLKNLNNEGMFYATPSASLEVIKYD-----LSYLKEALAL	294	
Db	260	NHKYKVNVELSENTKLLKA-----EKIFPDLLKKEQNYIDKLQELKCN	302	
QY	295	I---KAKTGADTKPELTRSFNEQAKGLGNDGDRSNYYDFLKGQVQEQVENACN-----	345	
Db	303	IQESKVLITYFTRLOSLKQDKDELVSLHQSKLNTYHNEIKGQKQLEHLSTRENEIT	362	
QY	346	-----LKTKYFGLDWKFNSL--IMLSEQK--VERDIKLIELYSKYNOLI-----QS	389	
Db	363	QFNQVLEKNQVF-----FNQDKLISSVQKQKPVIEEEK--RLYSEYNDLITKKEBLTKB	415	
QY	390	SSFDNEELAILK	401	

```

; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 215
; LENGTH: 2668
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-124-368A-215

Query Match      6.6%; Score 135.5; DB 7; Length 2668;
Best Local Similarity 22.2%; Pred. No. 0.11;
Matches 98; Conservative 82; Mismatches 148; Indels 113; Gaps 22;

QY 1 MCDLRKTKLIDKISSL--ELYKYSIFFRNYIENVAEDCLKNGLILESAAHNV-SEVELAR 57
Db 713 LCNL---ELEGKITDLOKELNK-----EVEENEALREEVILLSELKSLPSEVERLR 760

QY 58 LKVLQKNALLNCIIIS-----YRPHGIGYVLVKTDTLIDLEQPNIELPIGF 104
Db 761 KEIQDKSELHIITSEKDKLFSEVHVHKSERVOGLLEIGTKDDLTATQS-----NY 812

QY 105 EYLDVEY--VRDLGVDPDHITYKVKNNKNNSDAVKIHKSRLIYENFDY---ILKRYV 159
Db 813 KSTDOEQFNKTLHMDFEQYKMWLEENRNMQEIIVNLSKEA---QKFDSSLGALKTEL 868

QY 160 PCYTESFLDYLPEKIYVEIERRIENHNFYKDESIVLODALSSATTSLSALTQSN 219
Db 869 SYKTQE-----LQEKTR-EVOERL-----NEMEQLKEQLENRDLSTLQTV----- 906

QY 220 DRGSGILSSFLRK--QSNNSHSDISNLRNLDLSLQELARLKSNNNEGMYTATPSAS 277
Db 907 EREKTLITEKLOOTLEEVKTLTQEKDDLKQLESQIERDQKSDIHD-----TVANN 959

QY 278 L---EVIKYDLSYLK---EALALIKAKIGADTKPELTRSFNE-----QAKGLNDGKG 324
Db 960 IDTOEQLRNALLESKQHOETINTLKSISEVSRNLHMEENTGETKDFQKQWGVGIDKKQ 1019

QY 325 DRSNYDFLKGVOEQVENACNLKTYFGLDKMFNSLIMLSEQKVERDIKLIELYSKYN 384
Db 1020 D-----LEAKNTQTLTADVKNENII---EQOR--KIFSLIOEKNEIQ 1056

QY 385 QLIOSSSFDNEELAI-LKEKL 404
Db 1057 QMLESVIAEKEQKLTDLKENI 1077

RESULT 4
US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence

; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 214
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-124-368A-214

Query Match      6.6%; Score 135.5; DB 7; Length 2665;
Best Local Similarity 22.2%; Pred. No. 0.11;
Matches 98; Conservative 82; Mismatches 148; Indels 113; Gaps 22;

QY 1 MCDLRKTKLIDKISSL--ELYKYSIFFRNYIENVAEDCLKNGLILESAAHNV-SEVELAR 57
Db 713 LCNL---ELEGKITDLOKELNK-----EVEENEALREEVILLSELKSLPSEVERLR 760

QY 58 LKVLQKNALLNCIIIS-----YRPHGIGYVLVKTDTLIDLEQPNIELPIGF 104
Db 761 KEIQDKSELHIITSEKDKLFSEVHVHKSERVOGLLEIGTKDDLTATQS-----NY 812

QY 105 EYLDVEY--VRDLGVDPDHITYKVKNNKNNSDAVKIHKSRLIYENFDY---ILKRYV 159
Db 813 KSTDOEQFNKTLHMDFEQYKMWLEENRNMQEIIVNLSKEA---QKFDSSLGALKTEL 868

QY 160 PCYTESFLDYLPEKIYVEIERRIENHNFYKDESIVLODALSSATTSLSALTQSN 219
Db 869 SYKTQE-----LQEKTR-EVOERL-----NEMEQLKEQLENRDLSTLQTV----- 906

QY 220 DRGSGILSSFLRK--QSNNSHSDISNLRNLDLSLQELARLKSNNNEGMYTATPSAS 277
Db 907 EREKTLITEKLOOTLEEVKTLTQEKDDLKQLESQIERDQKSDIHD-----TVANN 959

QY 278 L---EVIKYDLSYLK---EALALIKAKIGADTKPELTRSFNE-----QAKGLNDGKG 324
Db 960 IDTOEQLRNALLESKQHOETINTLKSISEVSRNLHMEENTGETKDFQKQWGVGIDKKQ 1019

QY 325 DRSNYDFLKGVOEQVENACNLKTYFGLDKMFNSLIMLSEQKVERDIKLIELYSKYN 384
Db 1020 D-----LEAKNTQTLTADVKNENII---EQOR--KIFSLIOEKNEIQ 1056

QY 385 QLIOSSSFDNEELAI-LKEKL 404
Db 1057 QMLESVIAEKEQKLTDLKENI 1077

RESULT 3
US-11-124-368A-215
; Sequence 215, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin

```



```
Db 61 KLEKVPDVLMEYK-AIGGKIYI--VDGDITKH-ISLEALSEDKKKIK-----DIYGKDA 112
Qy 66 LNCIISYRFHGIGVYV--KTKDTLIDLQPVNLELPG-----FEYLDYEVVDLG 116
Db 113 LUHEHYVAKGYPVLVIQSSDVEVTEKALNVYVIGKILSDILSKINQPYQKFLD 172
Qy 117 V-----DFDH-----ITYKVKNNKNSLDVAKIHKSL--IYENFDY-----I 154
Db 173 VLNTIKNASDSQDQLLFTNQLKEHPTDFSVEFFLEQNSNEVQEPKAFAYIIEPQHRDV 232
Qy 155 LKRYVPCYTESP-LLDIYLFYKIVIEIRIENHNFYKDESLVOLQDALSSATTSLSA 213
Db 233 LQLYAP--EAFNYMDKFEQENLSLEE-----LKDQMLSRYEKWEKIKQHYQH 280
Qy 214 LTQSNDRSGSL-----SSFLRKQNSNNHSHK 240
Db 281 WSDSLSEGRGLLKKLQIPIEPKDDIHSLSQBEKELLKRIQIDSSDFL-----STEKE 336
Qy 241 DISNLR-NLNDLSOE-----LARKNSLN-----NEGMPFYTA 272
Db 337 FLKKLQIDIRDSLSBEKELLNRIOVDSNPLSEKEKEFLKLLDIOPIYDINQRLQDTG 396
Qy 273 -----TPSASLEVIK-----YDLSYLKEALAL--IKAKIGADTYKE 305
Db 397 GLIDSPSINLVRKQYKRDQIKNIDALLHQSIGSTLYNKIYLYENNINNLATATLGADLVD 456
Qy 306 PLTRS-----PNEQAKGLGNDGKGRNYDYDFLKGVEQ----- 339
Db 457 STDNTKINRGIFNEFKK--NFKYSISNY--MIVDINERPALDNERLKWRQLSPDTRA 511
Qy 340 --VENACNLKLTYPFLDMKFNLSMLBEEQKVERDIKLIELYSKYNQLOIQSSSPD--- 393
Db 512 GYLENG-KLLIQNRIGLEIKDVQIILQSEKEVIRIDAKVVP-KSKIDTKIQEAQLNINQE 569
Qy 394 -NEELAILK-EKLPSP 407
Db 570 MNKALGLPKYTKLITP 585

RESULT 7
US-10-821-234-975
; Sequence 975, Application US/10821234
; Publication No. US200502551141
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 975
; LENGTH: 989
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-975

Query Match 5.5%; Score 113.5; DB 6; Length 989;
Best Local Similarity 20.8%; Pred. No. 1.1;
Matches 98; Conservative 80; Mismatches 158; Indels 135; Gaps 22;

Qy 29 IENVAEDCLKNGLIILSAHNVSVELARLKVKQLKN-----ALLNCIISYRFHGIGVYL 82
Db 443 LQEIQLDQKR---LES-----SEARQLQVLELSRAELVCLNTEISENSSDLSQKL 494
Qy 83 VKTKD-----TLIDLEQPVNLELPIGEYLD-YEYVRDLGVDFDHITYKYVSKNNKNS 134
Db 495 KETOSKYBEAMKEVLSVQKMKLGL-VSPESMDNYSHFELNVTBEEI--NVLKQDLQNA 551
```

```
Qy 135 LDAVKIHKSL-----LIYENFDYILKRVVPCYTESFLLDIYLFYKIVYVEIERIENHN 188
Db 552 LEESERNKRVLEBEKLVREKGTVIKPPVEYEE-----MKSSYCSV---IENNN 600
Qy 189 ----FLFYK-----DESLVOLQDALSSATTSLSALTQSNDRGSGITLSSFLRKQNSNNHSHK 240
Db 601 KEKAFLEKYEQAEEIMKLDKTLK-----SQMTQASDEADEMKAMRMIDELN--K 652
Qy 241 DISNLRNL-----NDLSLQ--- 254
Db 653 QVSELSOLYKEAQAELEDYRKRSLEDVTAEYIHKABHEKMLQLTNVSVRAKAEDALSEM 712
Qy 255 -----ELARKNSLN-----NEGMPFYTATPSASLEVIKYDLSYLKEALALIKA 297
Db 713 SQYSKVLNLTQLKQLVDAQKENSVSITEHLQVITTLTAKEWEEKISNLUKHLASKEV 772
Qy 298 KIGADTKTEPLTRSFNEQAKGLGNDGKGRSNNYDYDFLKGVOQOVENACNLKLTYPFLDMK 357
Db 773 EVAKLEKQLL-----BEKAAMTDAMVPRSSYEKLGSSLESESVSLAS-KLKESVKEKEK 825
Qy 358 FNSLI--MLSEQKVERDIKLE--LYSK-----YNQLOIQSSSDFDNEELAILK 401
Db 826 VHSEVQIRSEVSQVREKENIQTLLKSKOEQVNEELLQKFOQAQEFELAE 876

RESULT 8
US-11-052-554A-121
; Sequence 121, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 121
; LENGTH: 1167
; TYPE: PRT
; ORGANISM: Helicobacter pylori J99
US-11-052-554A-121

Query Match 5.4%; Score 112; DB 7; Length 1167;
Best Local Similarity 21.6%; Pred. No. 1.7; Indels 114; Gaps 19;
Matches 95; Conservative 65; Mismatches 165;

Qy 30 ENVAEDCLKNGLIILSAHNVSVELARLKVKQLKN---ALLNCIISYRFHGIGVY--LVK 84
Db 626 ERLEKDVAKN---LESKSGNKNKMEAKSQANSQKDEIFALINKEANRARAIAAYANLKG 682
Qy 85 TKOTLIDLEQPVNLELPIGEYLDY-----VRDLGVDFDHIT 123
Db 683 IKRELSDKLENINKDLKDFSKSFD-EFKNGKNKDFSKAEETLKALKGSKVKGOLGINPEWIS 741
Qy 124 YKVKSNKNNKNSLDVAKIHKSLIYENFDYILKRVVPCYTESFLLDIYLFYKIVYVEIER 183
Db 742 ---KVENLNAALNEFFKNGK-----NKDFSKVTQAKSDLENISKDVLIINOK----ITDK 787
Qy 184 IENHNF-----FYKDESLVOLQDALSSATTSLSALTOS--NND 220
Db 788 VDLNQAQVSAKATGDFSGVEQALADLKNFQSKQLAQQAQKNEDFNTGKNSALYQSVKNG 847
Qy 221 RSGGILSFLRKQNSNNHSHKDISNLRNNDLSQELARLKSNLNEGMPFYTATPSASLEV 280
Db 848 VNGTLVGNGLSKABATTLTKNFSDIKK---ELNAKLGNFNNN--NNGLENSTEP----- 897
```

Qy 281 IKYDLSYLKEALAIKAKIGADTKEPLTRSFNEQAKGLGNDGKDRSNYYDFLKGQBOV 340
Db 898 -----IYTVAKKVKAKID-----RLDQIASGLGVQQA-----ASFLLKRDHKV 937
Qy 341 ENACNLKLTQ-----YFGLDMKFNSLIMLSEBQKVE-----RDKLIELYSKYNQLI 387
Db 938 DDLSKVLGSANHEPIYATID-DLGGPPPLKRDHKVDLDSKVLGSLREQKLTOKIDNLQAV 996
Qy 388 ---QSSSFDNBEALILKEK 403
Db 997 SEAKASHFDNLDQMDIKLK 1015

RESULT 9

US-11-155-288-20
; Sequence 20, Application US/11155288
; Publication No. US20060008468A1
; GENERAL INFORMATION:
; APPLICANT: Simard, Chih-Sheng
; TITLE OF INVENTION: COMBINATIONS OF TUMOR-ASSOCIATED
; ; TITLE OF INVENTION: ANTIGENS IN DIAGNOSTICS FOR VARIOUS TYPES OF CANCERS
; FILE REFERENCE: MANK 050A
; CURRENT APPLICATION NUMBER: US/11/155,288
; PRIOR FILING DATE: 2005-06-17
; PRIOR APPLICATION NUMBER: 60/580,969
; PRIOR FILING DATE: 2004-06-17
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-155-288-20

Query Match 5.4%; Score 111; DB 7; Length 976;

Best Local Similarity 23.3%; Pred. No. 1.6;
Matches 102; Conservative 52; Mismatches 161; Indels 122; Gaps 19;

Qy 12 KISSLELYKYISIFFRNYIENVAEDCLNGLILESAAHNVSEVELARKVL--KNALLNC 69
Db 406 KILTMLOKQS-----SELEMTK-----LTNNKEVELEELKVLGKETLL-- 447
Qy 70 IISYRFHGIGYVVKTKDLDLEOPVNIELPIGFEYLDYEVVRDLGVDFDHT--Y 124
Db 448 YENKOFEKIAELKGTQELIGLLOAREKE-----VHDLQLTAITTSQYY 495
Qy 125 KVKSNNKNSLDVAKIHSRLIYENFDYILKRYVPCYTESFLL-----DIYLFEKIY 177
Db 496 SKEVKDLKTELENEKLNKTELTSCHCNKLSLENKELTQETSDMTLELNKQOEDINNKKQE 555
Qy 178 VEIERRIEN-----HNFYPKDESIVQLQDALSSATTSLSALTQSNDRGSGILSPF 229
Db 556 ERMUKQIENLQETQLRNELEYVREELKQKRDVK-----CKLDKS----- 597
Qy 230 LKQNSNHSKDISNRLNDSLSQELARLKSNNNGMFYATATPSASLEVIKYDLSVLK 289
Db 598 --EENCNLRKQVENKNKYIEELOQENKALKKKGTA-----SKQLNVYEIKVNKLE 647
Qy 290 EALALIKAKIG--ADT--KEPLTRGFNEQ-----AKGLNGDGKDRSNYYDFLKGQV 337
Db 648 LELESAKQKFGBITTYQKEIEDKKISENLEVEKAKVIADAE-----VKLQ 696
Qy 338 EQVENACNLKTKYFGLDMKFNSLIMLSEBQKVERDIKLI-----LYSKYNLIQSS 390
Db 697 KEIDKRCQHKIAE-----MWALMEKHKHQYD-KIIEERDSELGLYKSEQSSSL 745
Qy 391 SPDNE-ELAILKEKLFS 406
Db 746 RASLELSNLSKAEELS 762

RESULT 10

US-11-077-550-112
; Sequence 112, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chadcock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 05/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-11-077-550-112

Query Match 5.2%; Score 107.5; DB 7; Length 888;

Best Local Similarity 20.4%; Pred. No. 2.5;
Matches 99; Conservative 65; Mismatches 146; Indels 175; Gaps 22;

Qy 6 KTKUIDKISSLELYKYISIFFRNYIENVAEDCLNGLILESAAHNVSEVELAR----- 57
Db 474 KNSFSDLSKNERIEYNT-QSNYIEN---DPPINELILDIDL--ISKIELPRENTESLTD 527
Qy 58 -----LKVLKQKALLNLCIISYRFH 76
Db 528 FNDVPVVEKQPAIKKIFTDENTIFQYLYSQTFPDLDIRDISLTSFDDALLFSNKVYGF 587
Qy 77 GIGYVLVKTKDLDLEOPVNIELPIGFEYLDYEVVRDLGVDFDHTYKVKSNKNNSLD 136
Db 588 SMDY--IKTANKVVEAG-----LFAGWVKQIVNDF-----VIEANKSNMTWD 626
Qy 137 AVKIHKSRLLI-----YEN-FDY-----ILKRYVP-----CYTESFLLDI 170
Db 627 --KIADISLIVPYIGLALNVGNETAAGNFENAFETAGASILLEFIPELLIPVVGAFPLES 684
Qy 171 YL--FEKIYVEIERRIENHNFYKDESIVQLQDALSSATTSLS-----SALTQSNDRGSG 224
Db 695 YIDNKNKIITIDNALTNRNEKSDMWGLIYVAQ--WLSTVNTQFYTIKEGMYKALNYQAA 743
Qy 225 1-----LSSFLRKQNSNHSKDISNRLNDSLSQELARLKSNNNGMFY---TA 272
Db 744 LEEIKRYNTYSEKESKNINIDFNDS--KLNEGINQADNINNFINNGSVSVLMKKM 801
Qy 273 TPSASLEVIKYDLSYLKEALALIKAKIGADTKEPLTRSFNEQAKGLGNDGKDRSNYYD- 331
Db 802 IPLAVEKLLDFDNLKKNLL-----NYIDE 826
Qy 332 ---FLKGQBOVENACNLKTKYFGLDMKFNSLIMLSEBQKVERDIKLIELYSKYNQLIQ 388
Db 827 NKLYLGSAEYKES-----KVNKYLKTIIMPFDLSIYTN-----DTILIEMFNKYNSEIL 875
Qy 389 SSSFD 393


```

Qy 121 HITVKNKNSLDAVKIHKSL-IYENFDYILKRYVPCYTESFLDYLPEKIVE 179
Db 151 BALQLLEBEEIOPQDNMKFKYKRMFEFYDR-----QKGFINYKYSQ-----INKPTVPT 200
Qy 180 IERRIENHFLFY-KDESLVOLQDALSATTLSALTQSNDRSGIL--SSFLRKQNSN 236
Db 201 IDDIILKSHLVSEYNEDEIVLE-----SYRTNSLRKI---NSNHGIDIRANSLFTEQELLN 252
Qy 237 NNSKDISNLRNINLSDELARLKSNNNEG-----MFTYTAPSASLEVIK 282
Db 253 IYSOELLNRGNL--AAASDIVELLALKNFGVYLDVMDLPGHSDLPKTIISRPSSIGLDR 310
Qy 283 YDLSYLKEALALIKAGADTKPELTRSFNEQANG--LGNDGKCDRSNYYDFLKGVOEQ 339
Db 311 WEMIKL-EAIMKYKYINNYSENPKD-LDQQLKDNFKLIIESKSEKSEIFSKUENL--- 365
Qy 340 VENACNLKLTQVGLDMKFNLSIMLSEBQKVERDIKLJELYSKY-----NOLIOSSSF 392
Db 366 --NVSLEIKTAFALGSGVINO-ALISKOGSYLTNLVIEQVKNRYQFLNOHLNPAIESDNN 422
Qy 393 DNEELAILKEKILFS 406
Db 423 FTDTTKIFHDSLEN 436

RESULT 14
US-10-055-877-147
; Sequence 147, Application US/10055877
; Publication No. US20050288241A1
; GENERAL INFORMATION:
; APPLICANT: DeCristofaro, Marc
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Miller, Charles
; APPLICANT: Tchernev, Velizar
; APPLICANT: Zhong, Mei
; APPLICANT: Anderson, David
; APPLICANT: Ballinger, Robert
; APPLICANT: Grylach, Valerie
; APPLICANT: Spytek, Kimberly
; APPLICANT: Kelluli, Luca
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Guo, Xiaojia
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Andrew, David
; APPLICANT: Mezes, Peter
; APPLICANT: Patturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Eisen, Andrew
; APPLICANT: Wolenc, Adam
; APPLICANT: Baumgartner, Jason
; APPLICANT: Shimkets, Richard
; APPLICANT: Gusev, Vladimir
; APPLICANT: Vernet, Corine
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Pena, Carol
; APPLICANT: Shenoy, Suresh
; APPLICANT: Li, Li
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ference
; TITLE OF INVENTION: Novel Polypeptides and Nucleic Acids Encoded Thereby
; FILE REFERENCE: 21402-251
; CURRENT APPLICATION NUMBER: US/10/055,877
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: 60/262,892
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263,598
; PRIOR FILING DATE: 2001-01-23
; PRIOR APPLICATION NUMBER: 60/263,799
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/264,117
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/264,139
; PRIOR FILING DATE: 2001-01-25

```

```

; PRIOR APPLICATION NUMBER: 60/264,478
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/263,351
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/272,870
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/275,990
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/275,927
; PRIOR FILING DATE: 2001-03-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 512
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 147
; LENGTH: 1732
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-055-877-147

Query Match 5.1%; Score 105.5; DB 6; Length 1732;
Best Local Similarity 20.7%; Pred. No. 8.1;
Matches 59; Conservative 50; Mismatches 115; Indels 61; Gaps 9;

Qy 131 KNSLDAVKIHKSLIYENFDYILKRYVPCYTESFLDYLPEKIVEIERRIENHFL 190
Db 559 KNSKELKDAHCQRKLAHQEFMEINERLTHTQK-----QKLARHVRDKEEVDLV 610
Qy 191 FYKDESLVOLQDALSATTLS-----SALTQSNDRSGILSSFLRKQNSNNHSDISN- 244
Db 611 MQAESLQELRAERAKKELEVHTEALIAESKDRK-----LREQ-SRHYSKOLENE 662
Qy 245 ---LRNLSLS-----QELARLKSNNNEGMYFTATPSASLEVIKYDLSYLKEAL 292
Db 663 LEGLKQKQISYSPGICSTEHQOEITKLKTDLEKKSIFYEEBISKREGTHASEIKNLKKE 722
Qy 293 ALIKAKIGADTKPELTRSFNEQANGLNDGKDSNYYDFLKGVOEQVENACNLKLTQYF 352
Db 723 HDSEGOQLALNKEIMV--LKDLEKTRRESOSEEPENEFKQYVERK----- 769
Qy 353 GLDMKFNLSIMLSEBQKVERDIKLJELYSKYNQILQSSSFNDEL 397
Db 770 -----VLLTEENK-----KLTSLELDKLTSLYESLSLNQHL 800

RESULT 15
US-10-793-626-2388
; Sequence 2388, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2388
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2388

Query Match 5.0%; Score 104; DB 6; Length 698;
Best Local Similarity 20.3%; Pred. No. 3.2;
Matches 93; Conservative 82; Mismatches 171; Indels 112; Gaps 28;

Qy 4 LRKTKLI---DKISSLELYKYSIFRNVIENVAEDCLKNGLILESAAHNVEVELARLKV 60

```

```

Db 78 LEDTKLTLQDDTTEKIELFK-KLEFKQLLIADIDTSSTN-----EEVIDKTFEIEODFQNV 131
QY 61 QLKALLNCIIISYRPHGIGYVLVTKDTLIDLEQPVNIELPIGFEYLDYEVV----RDLG 116
Db 132 DL-NDLNEAVIHFELEGNYL-----KDTI-----LKPGF-YTNHQHVINAEDV- 174
QY 117 VDFDHITYKVKSNKNKNSL-DAVKIHS--RL-IYEN--FDYILKRY----- 158
Db 175 KDYKHLVQWLEDKNTTKIVYDAKTYVSAHRLGINIEFDVMLASYIIDPSRIDDVK 234
QY 159 --VPCYTESFLD-IYLF---EKIYVEIERRIENHFLFYKDESLVQLQDALSSATTSL 212
Db 235 SVVSLYGQNYVKDNITIFGKGKHHIPBEPILNEH-----IASVTEAIAAVPTMK 285
QY 213 ALTQSNDRGSGI-----LSSFLRKQSNHNSKDISNLRNLNDSLSQELARLKSNL 263
Db 286 SOLEDYNO-----IELLKOLELPLARILSEMEIEIGIYTDINDLKEMEFIQKKLDVLI 341
QY 264 N-NEGMFYTATPSASLEVIKYDLSYLKEALALIKAKIGADT-----KEPL---TR 309
Db 342 HESAGEAFNINSPKQGVVLPETLQLP---VIKKTGYSTAVDVLEKLGHEHPIIDDL 398
QY 310 SFNEQAKGLGNDKGDRSNYDFLKV---QEQVENACNLKT---KYFGLDMKFNSL-I 362
Db 399 EYRQLAK-----LQSTYVEGLQKVISKDHRIHTRFNQTLAQTRLSIDPNLQNIPI 450
QY 363 MLSEEQKVERDIKLIELYSKYNQLIQSSSFDNEELAIL 400
Db 451 RLEGRKIRKAFK----PTSKDSVILSADYSQIELRVL 484

```

Search completed: January 24, 2006, 20:55:33
Job time : 7.51393 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 15.4404 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-4
Perfect score: 2062
Sequence: 1 MCDLRKTYLIDKISSELYK.....QSSFDNEELAILKEKLPSP 407

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.80:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1019.5	49.4	413	2 F70211	hypothetical prote
2	186.5	9.0	3394	2 T18501	hypothetical prote
3	150.5	7.3	1957	2 T38077	hypothetical coile
4	148	7.2	2269	2 T38677	rhoetry protein -
5	146	7.1	1827	2 A35694	cuti protein - fis
6	146	7.1	1828	2 T41455	cuti protein - fis
7	145.5	7.1	971	2 C82880	conserved hypothe
8	144.5	7.0	2166	2 G70163	hypothetical prote
9	143.5	7.0	1169	2 A64505	P115 homolog - Met
10	141	6.8	819	2 E70105	P115 protein homol
11	140.5	6.8	864	2 B90395	purine NTPase [imp
12	140	6.8	703	2 I64226	DNA helicase II (m
13	137.5	6.7	1127	2 T28317	ORF MSV156 hypothe
14	137.5	6.7	1979	2 T16222	hypothetical prote
15	136	6.6	3724	2 T18427	hypothetical prote
16	135.5	6.6	706	2 D90124	hypothetical prote
17	134	6.5	1939	2 T18372	repeat organellar
18	133.5	6.5	2663	1 S28261	centromere protein
19	133	6.5	546	2 A93600	conserved hypothe
20	132.5	6.4	1780	2 T17272	hypothetical prote
21	131.5	6.4	1738	2 T14867	interaptin - slime
22	130.5	6.3	1073	2 S14032	kinesin-related pr
23	130.5	6.3	1085	2 T38378	kinesin-like prote
24	129.5	6.3	365	2 JC2559	flagellin flic-1 -
25	129.5	6.3	488	2 AD0735	conserved hypothe
26	129	6.3	2910	2 C89910	DNA-directed RNA p
27	128.5	6.2	1009	2 C89910	hypothetical prote
28	128	6.2	1183	2 F90559	conserved hypothe
29	128	6.2	2010	2 B71616	phosphatase (acid

30	127.5	6.2	488	2 AG0621	probable bacteriop
31	127.5	6.2	689	2 T18494	hypothetical prote
32	126.5	6.1	2401	2 T28676	rhoetry protein -
33	126	6.1	611	2 D82881	zinc metalloprotei
34	126	6.1	980	2 E71606	hypothetical prote
35	126	6.1	1138	2 A82939	membrane nuclease
36	125	6.1	1365	2 T18419	hypothetical prote
37	125	6.1	1875	2 S38173	myosin-like protei
38	125	6.1	2819	2 T09080	probable chloroqui
39	124.5	6.0	944	2 S26710	spindle pole body
40	124.5	6.0	1021	2 E64576	hypothetical prote
41	124.5	6.0	1272	2 H82926	conserved hypothe
42	124.5	6.0	1294	2 T18473	hypothetical prote
43	124	6.0	1031	2 H81288	probable sugar tra
44	123.5	6.0	871	2 E97035	DNA polymerase I,
45	123	6.0	772	2 B82888	phenylalanine-trNA

ALIGNMENTS

RESULT 1

F70211
hypothetical protein BBA38 - Lyme disease spirochete plasmid A/lp54
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: F70211
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390: 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70211
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-413 <KLE>
A:Cross-references: UNIPROT:O50931; UNIPARC:UPI000005691F; GB:AE000790; NID:g2690224; P1
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid
C:Superfamily: Borrelia burgdorferi hypothetical protein BBA38

Query Match	49.4%;	Score	1019.5;	DB 2;	Length	413;			
Best Local Similarity	53.0%;	Pred. No.	1.2e-52;						
Matches	212;	Conservative	72;	Mismatches	99;	Indels	17;	Gaps	7;
QY	12	KISSELYKYSIPFRNYIENVAEDCLNGL---	ILESAAHVSEVELARLVQLKVALIN	68					
Db	20	KINPIDVYRYSIPFRNYIENVAEDCLNGL---	ILESAAHVSEVELARLVQLKVALIN	79					
QY	69	CIISVRFHGIGVAVKTKTLDLEQPNVIELPGFEYLDYEVYRDLGVDVDHITYKVS	128						
Db	80	AMISVRFNGAGYILVPKASEDELSKVNSELPFGFKLDFOKINKR-DSSYVEY--LS	136						
QY	129	NNKN-----NSLDAVKIHKSRLLIYENFDYILKRVPCYTESFLDIYLFEXIYVEIERI	184						
Db	137	NSKPDPPERARVVKIDKSRVIIYENVDYVLGEQEPAYQTSLNLLNICLEQIYLEIKRI	196						
QY	185	ENHNFIFYKDESLVQLQDALSSATTSALTQSNDRSGSLSSFLRKQNNHNSKDISN	244						
Db	197	RNYNLFYKDEHLVGLVESLEIAKEINVLA---NSKKG-IFSTFFKAEFN---KSPQA	249						
QY	245	LRNLDISIQLARLKNLNNEGFFYATPSASLEVIKYDLSYLKEALAKAKIGADTK	304						
Db	250	LSSVDELSRELSKIKNTLNNDGIFYTASENARLEVIKYDLEFLKDAFELVAKIGADTK	309						
QY	305	EPLTRSFNEQAKGLGNDGDRSNYYDFLKGQVQVENACNLKTYKFGDKMKNLSIML	364						
Db	310	EPLTRSFNEQVKGIGSSGKDGKSNYYDYLVKGQSVANACNLKLYRLLNKNFNEALEAL	369						
QY	.365	SEQKVERDIKILIELYSKYNQLIOSSSPDNBELAILKEKL	404						

I64226

DNA helicase II (mutB1) homolog - Mycoplasma genitalium
 C:Species: Mycoplasma genitalium
 C>Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 09-Jul-2004
 C:Accession: I64226; S18697
 R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
 M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
 , C.A.; Venter, J.C.
 Science 270, 397-403, 1995
 A:Title: The minimal gene complement of Mycoplasma genitalium.
 A:Reference number: A64200; MUID:96026346; PMID:7569993
 A:Accession: I64226
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-703 <TIG>
 A:Cross-references: UNIPROT:P47486; UNIPARC:UPI0000137EPD; GB:U39703; GB:L43967; NID:938
 A:Experimental source: strain G-37
 R:Peterson, S.N.; Schramm, N.; Hu, P.; Bott, K.F.; Hutchison III, C.A.
 Nucleic Acids Res. 19, 6027-6031, 1991
 A:Title: A random sequencing approach for placing markers on the physical map of Mycopla
 A:Reference number: S18693; MUID:92051396; PMID:1945886
 A:Accession: S18697
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 277-345 <PET>
 A:Cross-references: UNIPARC:UPI000016FA46; EMBL:X61517; NID:944318; PIDN:CAA43729.1; PID
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1991
 C:Genetics:
 A:Genetic code: SGC3
 C:Superfamily: helicase II

```

Query Match      6.8%; Score 140; DB 2; Length 703;
Best Local Similarity 21.3%; Pred. No. 0.79;
Matches 105; Conservative 73; Mismatches 140; Indels 176; Gaps 26;

Qy      4 LRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLKNGLILESAAHNSEVELARLVQV-- 61
Db      94 LKRRFIID-----DISYFLKEFLADSKLDINLAKQIIDNFKNFTADFEINKLDQDER 146

Qy      62 -----LKNALLNCIISYR-----FHGIGYVLVKT----- 85
Db      147 LISLCEHSLLNKDEYSTLTQTLINAFISYERKNKILNNKLDPHDL---LIKTCLNLSNDN 203

Qy      86 -----KOTLIDLEQPVNIELPIGPEYLDYEVVRDLGVDFDHIIVYKVSNNK---- 131
Db      204 DLLNQWSEQFQHLVDFQDTN-----QIQYELIKML-VTKKNFLVGDNNQMIYR 254

Qy      132 -----NNSLDVAK-----IHKSRLIITYENPDYILKRYVPCYTESFLDIYLFPEK----- 175
Db      255 MRGAVNGIITALKHDFNVFKSN-EFFINQYRCDQNILAVANQILKIMAYEKQVKTENK 313

Qy      176 -----IYVEIERIENH-NFLPYKDESIVQDALSSATTSLTSALTQSNDRG 222
Db      314 LLFSTLNSDKKPVYFOAE-SVENQANWTFNKIKALNQ-----TEKINFKD 357

Qy      223 SGILSFLRKONSNHSDKISNRLNDSLQ-ELARLKSNNLNEGMYFATPSASLEVI 281
Db      358 MAIL---FRK-----NRDITTWELIEADGTPIPKQKSYFN-----QLV 394

Qy      282 KYDLSYLKEALALIKAKIGADTKPELTRSFNBQAKGLGNDGK-----GDRSNYYDFLKGV 336
Db      395 K-----LQRLVIAISRTNLDIKRAL-----QALKWSNDLKLWKQSDKTNLFDFLWKS 444

Qy      337 Q-EQVENACNLKTYFGLDMFNSLIMLSEBQKVERDIKLIELYS--KYNQLIOSSSFD 393
Db      445 ELNQKNSHSLKATGY-----FNLLIKLAEDQOI--NLLFTFLPKLKVQDQTIEN---- 492

Qy      394 NEELAILKEKLPFSF 407
Db      493 -----LLWKKLTUF 501

```

RESULT 13

T28317

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
 C:Species: Melanoplus sanguinipes entomopoxvirus
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
 C:Accession: T28317
 R:Afonso, C.L.; Tullman, B.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
 J. Virol. 73, 533-552, 1999
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A:Reference number: 220484; MUID:99102612; PMID:9847359
 A:Accession: T28317
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1127 <AFO>
 A:Cross-references: UNIPROT:Q9YVT6; UNIPARC:UPI00000F6900; EMBL:AF063866; NID:94049647;
 C:Genetics:
 A:Note: MSV156

```

Query Match      6.7%; Score 137.5; DB 2; Length 1127;
Best Local Similarity 20.6%; Pred. No. 2;
Matches 104; Conservative 86; Mismatches 168; Indels 147; Gaps 26;

Qy      8 KLIDKISSLELYK-----YSIFPRNYIENVAEDCLKNGLILESAAHNSEVELAR 57
Db      407 KKIDDIKNNLQKLEESYKKIDEQTEYKKNKINKEYND-----IIEKNNLQKLEEN 460

Qy      58 LKVVQ-----LKNAL---LNCIISYRFHGIGYVLVKT---DTLIDLEQPV--NIETLPIGF 104
Db      461 KKIDEQTEYKKNKINKEYNDIIEKNNLQKLEENKKNINDKLTCLKNDIESNTEL---F 517

Qy      105 EYLDYEVYRD-----LGVDFDHITYKVSNNKNSLDAVAKIHKSRLLI-----IYEN-- 150
Db      518 NKLNISDFKDKSREIAKLNTYEQLRKDLLENINKTNEL--MKLSDNKLSSLEQLYDSKK 575

Qy      151 --PDYILKRYVPCYTESFLDIYLFPEK--IYVEIERR-----IENHNFIFY 192
Db      576 NILDGIDKIYNSLKEKNDKIDYFSNTEKFDIYVNIENKFGNLDISIINKINDQFKEY 635

Qy      193 -----KDESLVOLQALSSATTSLSALTQSNDRGSGILSSFLRKONSNHSDKISNL 245
Db      636 INSKIDSKNSELSTWFDIFNAKQIASIT--NN-----IENISNKIKDINEF 681

Qy      246 RNLDLSLSQELA-----RLKSNLNNEGMYFATPSASLEVIKYDLSYLKEALAL 294
Db      682 IISNEDSSKELLDIRKYKQKQDKIKDAMNTEVKSFENTLQKIDISIKSNINELTNAYDI 741

Qy      295 IKAKIGADTKPELTRSFNBQAKGLGNDGKDRSNYYDFLKGVOEVENACNLKTYFGL 354
Db      742 INTR-----ANDLDDKLNNYSEFPKNLYNNASDLDDTIQKNDEKVK-QLMEYLEK 791

Qy      355 D-----MKFNSLIMLSEBQKVERDI-----KLIELYSKYNQL-- 386
Db      792 KKNQSIENIDVNNFIKELIKFNN---TETNKSINELLTNDINDKIFKLYKELINKIST 847

Qy      387 -----IQSSSPD--NEELAILKEKL 404
Db      848 NLLKIYKNEIDNVNNEKLSIVIENL 872

```

RESULT 14

hypothetical protein PFB0145c - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
 C:Accession: C71622
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: C71622
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1979 <GAR>

A;Cross-references: UNIPROT:O96133; UNIPARC:UPI000007EA68; GB:AE001375; GB:AE001362; NID
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PF80145c

Query Match 6.7%; Score 137.5; DB 2; Length 1979;
Best Local Similarity 20.9%; Pred. No. 4.2;
Matches 92; Conservative 77; Mismatches 154; Indels 117; Gaps 23;
QY 24 PFRNYIENVAEDCLK---NGLILEAAHNVSEVELARLV-QLKN---ALLNCIIISYRFH 76
DB 979 PFKNSIEDKSHLKKKHEKOLLSDK-----KEIEBKNNKIKELNNDIKKLQDEILVYKKQ 1033
QY 77 G-----IGYVLV--KTKDTLIDLEQPVNIELPIGFVLDYEVYRDGLGVDFPHIYK 125
DB 1034 SNAQQVDHKKKSWILLKDKSEKIKENQINVE-----K 1068
QY 126 VKSNKNNSLDAVKIHKSLRIIYENFDYILKRYVPCYTESFLDYLILFEKIYVE----- 179
DB 1069 NEEDLKKKDDDEIRILNEELVKYITILYNLKK-----DPLLQNDLSSKIDINSLTINE 1122
QY 180 --IERRIENHNFYKDE-----SLVQLQDALSSATTSLSALTQSNN----- 219
DB 1123 GMCVDKIEEH-ILDYDEINKRSNLFQKNEICSLTTEVMELANNKKNELIENNNKLNIV 1181
QY 220 DRGSGILSFLRKQNS--NNHSKOISLRNLNDSLSQELARLKSNNLNNEGMYTATPSAS 277
DB 1182 DQKKLKKKQVEKQKKEIKLNKQITKCNQITDELNEBEVK-----LNNE-----N 1227
QY 278 LEVIKY--DLSY--LKEALALIKAKIGADTKPELTRSFNEQAKGL--GNDGKGRSNVY 330
DB 1228 IELITYSNDLANKPDMKENNMLMKLDENEDNIIKGMKIDDMKEIKYREDEKKNLNEI 1287
QY 331 DFLKGVOQVENACNLKLTGYFLGDMKFNLSIMLSEQKV---ERDIKLIELYSKYNQLI 387
DB 1288 NNLK---KKNEDMC-----IKYNNWNIKYGDICVKEMSLTYSKETSLEYEQIKVYKDE-- 1338
QY 388 QSSSPDNEELAILKEKLFSP 407
DB 1339 KCSQYDEIRFQ-YDEKCFQY 1357

RESULT 15

T18427
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18427
R;Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A;Reference number: 218935
A;Accession: T18427
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3724 <LAW>
A;Cross-references: UNIPROT:O77320; UNIPARC:UPI000007935D; EMBL:Z98547; NID:e1325376; PI
C;Genetics:
A;Introns: 307/1; 1545/2
A;Note: C0335c

Query Match 6.6%; Score 136; DB 2; Length 3724;
Best Local Similarity 23.0%; Pred. No. 12;
Matches 75; Conservative 58; Mismatches 125; Indels 68; Gaps 16;
QY 116 GVPDFHIYTKV---SNKNNSLDVAKIHKSLRIIYENFDYIL---KRYVPCYTESFLLD 169
DB 429 GINFDDVTKCKIIDDNNNENIDNNIYVENNKKLDQKSYDLFSSGKKNLILGVNNEGEEFN 488
QY 170 IYLPKIVYVEIERRIENHNFYKDESLVQLDALSSATTSLSALTQSNDRGSGILSSF 229
DB 489 BEVFENIEKEIQERKDEKN-----KNDKTI-----YNNNEEQTDLDLRN 527
QY 230 LRKQNNHNSKDISLRNLNDSLSQELARLKSNNLNNEG--FYTATPSASLEVIKY---- 283

DB 528 INKTESINNNNDNN--NINNKKFNFKIRTEHILNKESISKHKGSPSRDRKEIKLYTTN 585
QY 284 ---DLSY-LKEALALIKAK-----IGADTKEP-----LTRSFNEQAKGLGNDGKGRD 326
DB 586 KNEDSTPELKKELEIITNNKVNVEEDIIGSNEDDEYIHLKENLKEDANEYNNNDKENKN 645
QY 327 SNYYDFLKG---VOEQVENACNLK---TKYFGLDMKFNSL-----IMLSEEQKVERD 373
DB 646 NKTKEILKSKNYLNEKRTLEELKLRGKNNTFKKDEKYNLSGLGEVIINEIQIENEENKI-ND 704
QY 374 IKLIELYSKYNQLIOSSSFDNEELAI 399
DB 705 IQDGNI-SK-QKIIQSSSRRTNDTFNI 728

Search completed: January 24, 2006, 19:59:00
Job time : 17.4404 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 87.4555 Seconds
(without alignment)
3283.383 Million cell updates/sec

Title: US-10-688-058-4
Perfect score: 2062
Sequence: 1 MCDLRKTKLIDKISLSLEYK.....QSSSFDNEELAILKEKLFSP 407

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	2062	100.0	Q9S0G7_BORBU	Q9S0G7 borrelia bu
2	2051	99.5	Q9R3K2_BORBU	Q9R3K2 borrelia bu
3	2032	98.5	Q5XYN9_BORGA	Q5XYN9 borrelia ga
4	2019	97.9	Q9S0A2_BORBU	Q9S0A2 borrelia bu
5	2009.5	97.5	Q9S057_BORBU	Q9S057 borrelia bu
6	1894	91.9	Q5XZ87_BORGA	Q5XZ87 borrelia ga
7	1529.5	74.2	Q5XYI6_BORGA	Q5XYI6 borrelia ga
8	1175	57.0	Q9KKC1_BORHE	Q9KKC1 borrelia he
9	1019.5	49.4	Q5O931_BORBU	Q5O931 borrelia bu
10	664	32.2	Q5XZ76_BORGA	Q5XZ76 borrelia ga
11	614	29.8	Q6ASJ2_BORGA	Q6ASJ2 borrelia ga
12	189	9.2	Q5XYI5_BORGA	Q5XYI5 borrelia ga
13	186.5	9.0	Q77384_PLAF7	Q77384 plasmidium
14	181.5	8.8	Q9BK46_PLAFA	Q9BK46 plasmidium
15	181.5	8.8	Q8IDX6_PLAF7	Q8IDX6 plasmidium
16	181.5	8.8	Q7YWE9_PLAF7	Q7YWE9 plasmidium
17	181.5	8.8	Q9BK45_PLAFA	Q9BK45 plasmidium
18	181	8.8	Q7YWF0_PLAFA	Q7YWF0 plasmidium
19	181	8.8	Q7YWF1_PLAFA	Q7YWF1 plasmidium
20	181	8.8	Q7YWE6_PLAFA	Q7YWE6 plasmidium
21	181	8.8	Q7YWE7_PLAFA	Q7YWE7 plasmidium
22	175	8.5	Q8I462_PLAF7	Q8I462 plasmidium
23	166	8.1	Q6L0R1_PICTO	Q6L0R1 picophilus
24	165.5	8.0	Q7YWF2_PLAFA	Q7YWF2 plasmidium
25	165.5	8.0	Q7YWE8_PLAFA	Q7YWE8 plasmidium
26	163	7.9	Q7YWE3_PLAFA	Q7YWE3 plasmidium
27	161	7.8	Q8EWP8_MYCPE	Q8EWP8 mycoplasma
28	160.5	7.8	Q8CPC5_STAEP	Q8CPC5 staphylococ
29	159.5	7.7	Q8IIL3_PLAF7	Q8IIL3 plasmidium
30	159	7.7	Q8I4V4_PLAF7	Q8I4V4 plasmidium
31	157.5	7.6	Q5HPJ3_STAEP	Q5HPJ3 staphylococ

32	157.5	7.6	1455	2	Q8IKG8_PLAF7	Q8IKG8 plasmidium
33	157	7.6	443	2	Q6FA80_ACID	Q6FA80 acinetobact
34	156	7.6	855	2	Q54QB4_DICDI	Q54QB4 dictyosteli
35	155.5	7.5	2093	2	Q7RJA2_PLAYO	Q7RJA2 plasmidium
36	155.5	7.5	5687	2	Q8IJH4_PLAF7	Q8IJH4 plasmidium
37	155	7.5	693	2	Q7RPY2_PLAYO	Q7RPY2 plasmidium
38	155	7.5	915	2	Q4UGA6_THEAN	Q4UGA6 theileria a
39	155	7.5	928	2	Q8IJN2_PLAF7	Q8IJN2 plasmidium
40	154	7.5	861	2	Q8IHT6_PLAF7	Q8IHT6 plasmidium
41	154	7.5	895	2	Q7RBZ1_PLAYO	Q7RBZ1 plasmidium
42	154	7.5	1232	2	Q8IEF5_PLAF7	Q8IEF5 plasmidium
43	153	7.4	1272	2	Q8I480_PLAF7	Q8I480 plasmidium
44	152.5	7.4	1241	2	Q5CT95_CRYPV	Q5CT95 cryptospori
45	150.5	7.3	1957	1	SPO15_SCHPO	Q10411 schizosacch

ALIGNMENTS

RESULT 1

Q9S0G7_BORBU
ID Q9S0G7 BORBU PRELIMINARY; PRT; 407 AA.
AC Q9S0G7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein BBR01.
GN OrderedLocusNames=BBR01;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-4.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=1139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P., Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Hatt D.H., Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi.";
RT Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K., Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D., Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J., Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A., Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K., Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AE001577; AAF07516.1; --; Genomic_DNA.
DR TIGR; BBR01; --;
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 407 AA; 47004 MW; 752CC25E173B8781 CRC64;

Query Match 100.0%; Score 2062; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 2.4e-109;
Matches 407; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MCDLRKTKLIDKISLSLEYKSIFFRNYIENVAEDCLKNGLILSAAHNVSEVELARLKV 60

Db 1 MCDLRKTKLIDKISLSLEYKSIFFRNYIENVAEDCLKNGLILSAAHNVSEVELARLKV 60


```
Matches 400; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLNGLILSAHNVSSEVELARLKV 60
DB 1 MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLNGLILSAHNVSSEVELARLKV 60
QY 61 QKALLNCIIISYRPHGIGYLVKTKDTLIDLEQPNVIELPIGFYLDYVYRDIGVDFD 120
DB 61 QKALLNCIIISYRPHGIGYLVKTKDTLIDLEQPNVIELPIGFYLDYVYRDIGVDFD 120
QY 121 HITKVKNNKNSLDVAKIHKSRLLIYENFDYILKRYVPCYTESFLDIIYLFKIIYVEI 180
DB 121 HITKVKNNKNSLDVAKIHKSRLLIYENFDYILKRYVPCYTESFLDIIYLFKIIYVEI 180
QY 181 ERIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
DB 181 ERIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
QY 181 EGRIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
DB 181 EGRIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
QY 241 DISNLRNLSLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300
DB 241 DISNLRNLSLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300
QY 301 ADTKEPLTRSFNEQAKGLGNDGKDRSNYYDFLKGVOEQVENACNLKTYFGLDMKFN 360
DB 301 ADTKEPLTRSFNEQAKGLGNDGKDRSNYYDFLKGVOEQVENACNLKTYFGLDMKFN 360
QY 361 LIMLSEKQVERDIKLIELYKYNQIIOSSSFDNEELAILKEKLFSP 407
DB 361 LIMLSEKQVERDIKLIELYKYNQIIOSSSFDNEELAILKEKLFSP 407

RESULT 4
Q9S0A2 BORBU PRELIMINARY; PRT; 407 AA.
AC Q9S0A2;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE Hypothetical protein BBO01.
GN OrderedLocNames=BBO01;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi."
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
DR EMBL; AF001579; AAF07602.1; -; Genomic_DNA.
DR TIGR; BBO01; -;
```

```
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 407 AA; 47061 MW; 8E34A6BF392D15EE CRC64;

Query Match 97.9%; Score 2019; DB 2; Length 407;
Best Local Similarity 97.5%; Pred. No. 6,6e-107;
Matches 397; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 1 MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLNGLILSAHNVSSEVELARLKV 60
DB 1 MCDLRKTKLIDKISSLELYKYSIFPRNYIENVAEDCLNGLILSAHNVSSEVELARLKV 60
QY 61 QKALLNCIIISYRPHGIGYLVKTKDTLIDLEQPNVIELPIGFYLDYVYRDIGVDFD 120
DB 61 QKALLNCIIISYRPHGIGYLVKTKDTLIDLEQPNVIELPIGFYLDYVYRDIGVDFD 120
QY 121 HITKVKNNKNSLDVAKIHKSRLLIYENFDYILKRYVPCYTESFLDIIYLFKIIYVEI 180
DB 121 HITKVKNNKNSLDVAKIHKSRLLIYENFDYILKRYVPCYTESFLDIIYLFKIIYVEI 180
QY 181 ERIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
DB 181 ERIENHNFLFYKDESLVQLDALSSATTSLSALTQGNDRSGILSSFLRKQNSNHSK 240
QY 241 DISNLRNLSLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300
DB 241 DISNLRNLSLSQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAKIG 300
QY 301 ADTKEPLTRSFNEQAKGLGNDGKDRSNYYDFLKGVOEQVENACNLKTYFGLDMKFN 360
DB 301 ADTKEPLTRSFNEQAKGLGNDGKDRSNYYDFLKGVOEQVENACNLKTYFGLDMKFN 360
QY 361 LIMLSEKQVERDIKLIELYKYNQIIOSSSFDNEELAILKEKLFSP 407
DB 361 LIMLSEKQVERDIKLIELYKYNQIIOSSSFDNEELAILKEKLFSP 407

RESULT 5
Q9S057 BORBU PRELIMINARY; PRT; 408 AA.
AC Q9S057;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE Hypothetical protein BBN01.
GN OrderedLocNames=BBN01;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi."
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
```



```
Db 1 MCNLRKTKLVDRINLLKLYRYSIFRNVIENVAEDCLNGLVLESTSHNVSEFELDRKV 60
QY 61 QLKNALNCIIISYRFH--GIGYVLVKTOTLDLEQPVNIELPIGFEYLDYEVYRDGLVD 118
Db 61 QLKNALNCIIISYRFHNGIGYVLVKTQDALLDEPNIELPIGFEYLDYEVSKDSIGD 120
QY 119 FPHITYKVKNNKNSLDAVKHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYV 178
Db 121 FDIYINVKPNKNGSNFNTVKHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYV 180
QY 179 ETIERIENHFLPYKDESIVOLQDASSATTSLALTO--SNNDRSGILSSFLRKNSNN 237
Db 181 ETIERIENHFLPYKDESIVOLQDASSATTSLALTO--SNNDRSGILSSFLRKNSNN 240
QY 238 HSKDLSNLRNLSLQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIK 297
Db 241 HSKDLSNLRNLSLQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIK 300
QY 298 KIGADTKPELTSFNEQAGLNGDKGDRSNYYDFLKGVOE 338
Db 301 KIGADTKPELTSFNEQAGLNGDKGDRSNYYDFLKGVOE 341

RESULT 8
Q9KKC1 BORHE
ID Q9KKC1 BORHE PRELIMINARY; PRT; 404 AA.
AC Q9KKC1;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE Hypothetical protein.
OS Borrelia hermsii.
OG Plasmid cp32.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=140;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-1;
RX MEDLINE=20316012; PubMed=10858201; DOI=10.1128/JAI.68.7.3900-3908.2000;
RA Stevenson B., Porcella S.F., Oie K.L., Fitzpatrick C.A., Raffel S.J., Lubke L., Schrupp M.E., Schwan T.G.;
RT "The relapsing fever spirochete Borrelia hermsii contains multiple, antigen-encoding circular plasmids that are homologous to the cp32 plasmids of Lyme disease spirochetes."
RL Infect. Immun. 68:3900-3908(2000).
DR EMBL; AF123078; AAF28886.1; -; Genomic_DNA.
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 404 AA; 46409 MW; 5567119DBB024671 CRC64;

Query Match 57.0%; Score 1175; DB 2; Length 404;
Best Local Similarity 57.9%; Pred. No. 6.4e-59;
Matches 231; Conservative 69; Mismatches 89; Indels 10; Gaps 3;

QY 13 ISSLELYKYSIFRNVIENVAEDCLNGLILSAHNSEVELARLKVQLKNALNCIIIS 72
Db 7 INSRDLYKYSIFRNVIENVAEDCLNGLILSAHNSEVELARLKVQLKNALNCIIIS 66
QY 73 YRPHGIGYVLVKTOTLDLEQPVNIELPIGFEYLDYEVYRDGLVDYFPHITYKVKNSN-- 130
Db 67 YRPHGAGYILVKTNDNFSLSESUNLELPFGFYLDYSRVYDAGPGSNCTIYYAGEDAF 126
QY 131 KNSLDAVKHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYVIERIENHNL 190
Db 127 DNLSELSLHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYVIERIENHNL 186
QY 191 FYKDESIVOLQDASSATTSLALTO--SNNDRSGILSSFLRKNSNNHSDIENLRN 247
Db 187 FYKDESLSLHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYVIERIENHNL 241
QY 248 LNDLSLQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAGADTK 307
```

```
Db 242 ANDELMRELEKSLKNDNDGIFYSCEPNASLEVIKYDLSYLKEALSIVKAKIGADTK 301
QY 308 TRSNEQAKGLNGDKGDRSNYYDFLKGVOQVQVENACNLKLTIFYGLDMKFNLSIMLSEE 367
Db 302 TRSNEQAKGLNGDKGDRSNYYDFLKGVOQVQVENACNLKLTIFYGLDMKFNLSIMLSEE 361
QY 368 QKVERDKLIELYSKYNTQLOSSFDNEELAILKEKLF 406
Db 362 EKIRDMRLLEIYERYSSVISNPNSLDERMKLKENLFT 400

RESULT 9
O50931 BORBU PRELIMINARY; PRT; 413 AA.
AC O50931;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DE Hypothetical protein BBA38.
GN OrderedLocustNames=BBA38;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K., Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D., Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J., Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A., Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K., Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."
RT Nature 390:580-586(1997).
DR EMBL; AE000790; AAC66300.1; -; Genomic_DNA.
DR PIR; P70211; P70211.
DR TIGR; BBA38; -.
DR InterPro; IPR009427; DUF1073.
DR Pfam; PF06381; DUF1073; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 413 AA; 47901 MW; 382654B2DF351241 CRC64;

Query Match 49.4%; Score 1019.5; DB 2; Length 413;
Best Local Similarity 53.0%; Pred. No. 4.5e-50;
Matches 212; Conservative 72; Mismatches 99; Indels 17; Gaps 7;

QY 12 KISSLELYKYSIFRNVIENVAEDCLNGL--ILESAHNSEVELARLKVQLKNALIN 68
Db 20 KINPIDVTRYSIFRNVIENVAEDCLNGLINLSLEKSLKSSLSLGLKVQLKEALLE 79
QY 69 CIISYRPHGIGYVLVKTOTLDLEQPVNIELPIGFEYLDYEVYRDGLVDYFPHITYKVKNS 128
Db 80 AMISYRPHGAGYILVVKPKSEDELSKVNSELPTGFKYLDQKINKR--DSSYVEY--LS 136
QY 129 NNKN-----NSLDVAKHKSRLLIYENFDYILKRYVPCYTESFLDDIYLPKIIYVIERI 184
Db 137 NSKDPDFERARVVVKIDKSRVVIYENYDVLGEQEPAYTQTSLLNLCLEQIYLEIKRI 196
QY 185 ENHNFIFYKDESIVOLQDASSATTSLALTO--SNNDRSGILSSFLRKNSNNHSDIEN 244
Db 197 RYNFNFIFYKDEHVLGVLESLEAKEEINVL--NSKGK--IFSTFFKQEPN---KSPQA 249
QY 245 LNLNLSLSLQELARLKSNNLNNEGMYTATPSASLEVIKYDLSYLKEALALIKAGADTK 304
Db 250 LSSVDELSELRLSKNTLNDCIYTTASENARLEVIKYDLEFLKDAPELVKAKIGADTK 309
```


O77384 PLAF7
ID O77384_PLAF7 PRELIMINARY; PRT; 3394 AA.
AC O77384;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein PF0760C
GN Name=PF0760C; Synonyms=MA13P6.11;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=9376885; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kves S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum.";
RL Nature 400:532-538(1999).
DR EMBL: Z98551; CAB1140.1; -; Genomic_DNA.
DR PIR: T18501; T18501.
KW Hypothetical protein.
SQ SEQUENCE 3394 AA; 402949 MW; 81BB86F2056A4B57 CRC64;
Query Match 9.0%; Score 186.5; DB 2; Length 3394;
Best Local Similarity 21.5%; Pred. No. 0.13;
Matches 104; Conservative 89; Mismatches 181; Indels 109; Gaps 20;
QY 1 MCDLRTKLIDKISSLELYKYIFRNVIENVAEDCLKNGILLESAAHNVSEVLARLV 60
Db 1561 MCDVTLASSGVNKKDD-----FLNLEENEENSFLEVEIRIKSLQELCDKSEILKI 1612
QY 61 Q-LKVALNCIISYRFGHIGYVLTQTLIDLEQVNIELPIGFEYL-----DYEV 111
Db 1613 KGEKNILITCETWK-----CFCKNSKEBISRLKEICKEQLEKHEFLINKSNEDKLKY 1667
QY 112 VRDLGVD----FDHITYKVKSNKNSLDV-----KHKSRLLIYENFDYILKRYVPC 161
Db 1668 INSLCDEKDYDVVVDKIK-NMREIDKLNNDINEXYEIKLKHNNLNINEMNLIK 1726
QY 162 YTESFLDLTYLPEKIYVETERR-----IEN-HNPLFVKDESLOLODSSATTSLSALTO 216
Db 1727 NKETENMIAKQEEYDIKIKDKTKNIQNEYNDLLEKNEVVVKNMLYN----DMNVLLK 1783
QY 217 SNDRGSGILSFLRKQNSNNHSDISNRLN-----NDLSLQELARLKSNNNEG 267
Db 1784 EHKE-----EFLLEKENIKILQKQNTVNDMFKNQINVDNLLKNRLDQL-FNINQDL 1836
QY 268 MEYATPASPASLEVIKYDLSYLKEALALIKAGIDTK-----EPLTRSFNQAKGL 318
Db 1837 QKHLDTNQKHLQKLYDYIEIKERLIKTEKTKINKQEKYIIQKQDNLLINDFNSTTTT 1896
QY 319 GNDGKGRSN-----YYDFLKGVOEVENACNLKTKYFGLDMKFNLSLMS 365
Db 1897 NNNNNNDNNNDNNNDNNNDNTYQOFIHLKANLENS-RLELKELSN-----NEKIQLS 1951
QY 366 EQQ-----KVERD-IKLEL-----YSKYNQLIQSSFDNBEALIL 400
Db 1952 DEKNRMKITLEDKLFKNEKQKWKQLQIIDNNKNKVMIQNKLTNLDMLSEENMLLN 2011
QY 401 KEK 403
Db 2012 KEE 2014
RESULT 14
Q9BK46_PLAF6
ID Q9BK46_PLAF6 PRELIMINARY; PRT; 3130 AA.

O9BK46;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Reticulocyte binding protein 2 homolog A.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21101060; PubMed=11160005;
RX DOI=10.1128/IAI.69.2.1084-1092.2001;
RA Triglia T., Thompson J., Caruana S.R., Delorenzi M., Speed T.,
RA Cowman A.F.;
RT "Identification of proteins from Plasmodium falciparum that are
RT homologous to reticulocyte binding proteins in Plasmodium vivax.";
RL Infect. Immun. 69:1084-1092(2001).
DR EMBL: AF312916; AAK19244.1; -; Genomic DNA.
SQ SEQUENCE 3130 AA; 370419 MW; 13D973DB89D82026 CRC64;
Query Match 8.8%; Score 181.5; DB 2; Length 3130;
Best Local Similarity 19.2%; Pred. No. 0.22;
Matches 95; Conservative 111; Mismatches 137; Indels 153; Gaps 23;
QY 11 DKISSLELYKYIFRNVIENVAEDCLKNGILLESAA-----HNSEVELARLVQV-LKNA 65
Db 535 EKASWDEMEKKELSLNNYIEK-TDYILQTYNIFKSKNIIINNNSKNISKYITIEGLKND 593
QY 66 L-LNCIISYRFGHIGYVLTQTLI---DLEQVNIELPIGFEYL---YEVVRLDGLVD 118
Db 594 IDELSLSISV-----FKDSQETLIKDDKKNMKTDLNNVYKIENVTINEIILL 645
QY 119 FDHITYKVKSNKNSLDVAKHKSRLLIYENFDYILKRYVPCYTESFLDLTYLPEKIYV 178
Db 646 KDSITQRIADIDELNSLNIN-----DFINEKNISQEKVSYNLN-KLYKGSFE 694
QY 179 EIERRIENH---NPLFYKDESLOLODSSATTSLSALTOSNDRGSGILSFLRKONS 235
Db 695 ELESLSHFLPTKLYFLFEKKSQVNELOTLINTSNNECAK-----NFMKSDNN 741
QY 236 NNHKS-----DISNRLNDSLSOELA-RLKSNNNEGMYFTAPS----- 275
Db 742 NNNNSNIIINLKTLSHLLSKENIKKGLNHLIEQIONSNTYITTDINRMEDYK 801
QY 276 ---ASLEVIKYDLSYL-----KEALALIKAKIGADTK 304
Db 802 EEISLEVYKKTIGNIQKEYILHLVENDKNALAVHTSMQILOKDAIQNKIKSDDIK 861
QY 305 EPLTRSFNQAKGLGNDGKGRSNYDPL-KGVE-----QVENACNLKTKYF----- 352
Db 862 --ILKKYKEMNQ-----DLLNYYEILDKKLDKNTYIKEMHTASLVQITQIYIPEDKT 911
QY 353 --GLDMKENSIMLSEKQVE---RDIKLIIEYSK-----YNO 385
Db 912 ISELEQEPN-----NNKQDLNIIQDINAMNLNIIQTLNIGINACNTNNNNVEHLNKK 966
QY 386 LIQSSFDNEELAILK 401
Db 967 KIELKNILNDQMKIK 982
RESULT 15
Q9BK46_PLAF7
ID Q9BK46_PLAF7 PRELIMINARY; PRT; 3130 AA.
AC Q9BK46;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 10-MAY-2005 (TREMBLrel. 30, Last annotation update)
DE Reticulocyte binding protein 2 homolog a (Normocyte binding protein
DE 2a).
GN Name=PF13_0198;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	157.5	7.6	583	2	US-09-710-279-1358	Sequence 1358, Ap
2	157.5	7.6	1010	2	US-09-134-001C-5178	Sequence 5178, Ap
3	153.5	7.4	448	2	US-09-328-352-6856	Sequence 6856, Ap
4	136	6.6	2662	2	US-09-595-684B-31	Sequence 31, Appl
5	133.5	6.5	2663	2	US-09-538-092-1252	Sequence 1252, Ap
6	132.5	6.4	1780	2	US-09-949-016-6899	Sequence 6899, Ap
7	132.5	6.4	1786	2	US-09-949-016-7880	Sequence 7880, Ap
8	132	6.4	511	2	US-09-248-796A-16380	Sequence 16380, A
9	131.5	6.4	878	2	US-09-134-001C-4378	Sequence 4378, Ap
10	130.5	6.3	1073	2	US-09-541-783-6	Sequence 6, Appl
11	130.5	6.3	1073	2	US-09-723-820-6	Sequence 6, Appl
12	130.5	6.3	1073	2	US-10-270-085-6	Sequence 6, Appl
13	129.5	6.3	575	2	US-09-724-519-8	Sequence 8, Appl
14	129.5	6.3	575	2	US-09-592-037-8	Sequence 8, Appl
15	129.5	6.3	575	2	US-09-428-156B-8	Sequence 8, Appl
16	129.5	6.3	905	2	US-09-248-796A-16333	Sequence 16333, A
17	129	6.3	1241	2	US-09-248-796A-18397	Sequence 18397, A
18	127.5	6.2	367	2	US-09-543-681A-8041	Sequence 8041, Ap
19	126	6.1	5024	2	US-09-710-279-2964	Sequence 2964, Ap
20	125.5	6.1	435	2	US-09-762-194-8	Sequence 8, Appl
21	125.5	6.1	435	2	US-09-762-194-10	Sequence 10, Appl
22	123.5	6.0	415	2	US-09-543-681A-8002	Sequence 8002, Ap
23	123	6.0	630	2	US-09-248-796A-20275	Sequence 20275, A
24	123	6.0	3878	2	US-09-914-259-11	Sequence 11, Appl
25	121	5.9	872	2	US-08-851-843A-8	Sequence 8, Appl
26	121	5.9	872	2	US-08-851-843A-54	Sequence 54, Appl
27	121	5.9	872	2	US-08-974-549A-221	Sequence 221, App

RESULT 2
US-09-134-001C-5178
; Sequence 5178, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5178
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5178

Query Match	7.6%	Score 157.5	DB 2	Length 1010
Best Local Similarity	25.6%	Pred. No. 9.9e-06		
Matches	80	Conservative 55	Mismatches 98	Indels 79
			Gaps	16
Qy	131	KNNSLDAVKHKSLIIYENFDYILKRVVPCYTSFLLDILFLFEKIYVEIERR	---	ISN- 186
Db	162	QSNKQKQSIILRT-LFNSERFDEI	-----RHLLVENVKQEK--VOIENRYTQIENL	209
Qy	187	-HNFPLYKDELSVOLQDALSAATTSLSALTQSNDRGSGIILSSFLRKQNSNNHKSIDISNL		245
Db	210	WNDIDTFNNDELALYKELESSQTDKMIETFPQFNDYGCKILKSP--EAKNKIYKELDDL		267
Qy	246	RN----LNDLSGQELARLKSNI--NNEGMFYATPASALESVIKYD-----LSYIKKALAL		294
Db	268	NHKYKVNVLSENTKKLKA-----EKIFDDLKQEQNYIDKLQKLM		310
Qy	295	I---KAKIGADTKPELFTSFNEQAKGLGNDGKDBSNYYDFLKGVOEQVENACN-----		345
Db	311	IQESKVLITYTFRLOSKKKQKDELVSLEHQSKJNETNYHNRIKGFQKQLEHLSTRENBIT		370
Qy	346	-----LKLTKYFGLDMKFNSL--IWLSEEQK--VERDIKTELKSKYNOLI-----QS		389
Db	371	QFNQYLEKNQVFP-----FNQLDKLISSYQQRPVIEEEK--RLYSEYNDLITKKEELTKE		423
Qy	390	SSFONEELAILK		401
Db	424	MNNKNDKFAITE		435

```

RESULT 3
US-09-328-352-6856
; Sequence 6856, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6856
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6856

Query Match          7.4%; Score 153.5; DB 2; Length 448;
Best Local Similarity 21.8%; Pred. No. 6.6e-06;

```

Matches	87;	Conservative	59;	Mismatches	145;	Indels	109;	Gaps	18;
Qy	26	RNYIENVAEDCLXNGLILESAA--HNYSEVELARLKVLQKNALLNCIISYRFHGIGYVLVK	84						
Db	59	KRIVNRPASDMLRAGWFYEGIQDRDLIKLEEACAKAQLNQLILLSRLYSRLYGWYVLLG	118						
Qy	85	TKDTLIDLEQVNI--ELPIG----FEYLDYEVYRDVGVPDHDIT-----	123						
Db	119	TVDG--GNLEQFPDLKGLGVGRLEFFTVLKKYIKP-----DTTYLSPKVCGGLLKQPEF	172						
Qy	124	YVKSSNNKNSLDAVKIHKSRLLIYENFDYILKRYVPCYTESFLDLDIYLPEKIYVEIERR	183						
Db	173	YKLQMDGKAPQ-----RIHHSRLIKFCHADVNEE-----PQSLQEVY-----	211						
Qy	184	IENHNFLYKDESIVQLQDALSSATTSLSALTOSNDRSGSILSSFLRKQNSNNHSHK--DI	242						
Db	212	-----EDLLDHAAYK-----KGSASLV-----HESKIDV	235						
Qy	243	SNLRNLNDSLQSLARLKNLNNEGFFYTATPSASLEVTK-----YDLSYLKEALALIK	296						
Db	236	IKTFLGLVDRIKEDMKAVAEFFLSVGLLKLGLNGMIVLDAEEDYDSTKYTNFAGLFDMMREFS	295						
Qy	297	AKTGADTKPELTPRSFNEQAKGLNDGKGRSRYD-----FLKGVQEQV--ENACNLK	347						
Db	296	IQTAGAADIPTTTLFGQSPAGMNATGSHDTRNTYDSIATKQTWMLKPFMMQILDVICQTT	355						
Qy	348	LTKYF--GLDMKFNLSIML-----SEQKV--ERDIKLIEL	379						
Db	356	FGRVFPNLDIVFNPQLQWLQDAKVRSEVEKANSERDAKYLEM	395						
RESULT 4									
US-09-595-684B-31									
; Sequence 31, Application US/09595684B									
; Patent No. 6544766									
; GENERAL INFORMATION:									
; APPLICANT: Beraud, Christophe									
; APPLICANT: Ohashi, Cara									
; APPLICANT: Sakowicz, Roman									
; APPLICANT: Vaisberg, Eugenii									
; APPLICANT: Wood, Kenneth									
; APPLICANT: Yu, Ming									
; TITLE OF INVENTION: Human kinesins and methods of producing									
; TITLE OF INVENTION: and purifying human kinesins									
; FILE REFERENCE: cytop036									
; CURRENT APPLICATION NUMBER: US/09/595,684B									
; CURRENT FILING DATE: 2002-06-24									
; PRIOR APPLICATION NUMBER: 09/295,612									
; PRIOR FILING DATE: 2000-04-20									
; NUMBER OF SEQ ID NOS: 105									
; SOFTWARE: FastSeq for Windows Version 4.0									
; SEQ ID NO 31									
; LENGTH: 2662									
; TYPE: PRT									
; ORGANISM: Human									
; US-09-595-684B-31									

Query Match	6.6%	Score 136;	DB 2;	Length 2662;
Best Local Similarity	22.1%;	Pred. No. 0.005;		
Matches	97;	Conservative	80;	Mismatches 153; Indels 108; Gaps 21;
Qy	1	MCRLRKTGLDKLSSL---	ELYKYSIFRNVIENVAEDCLKNGHLLSSAAHV--	SEVELAR 57
Db	713	LCNL---	ELEGKITDQKELNK-----	EVEENEAREFVILLSELKSLPSEVERLR 760
Qy	58	LKVQLKNALNCIIS---	---RFHGIGYVLVKTKDLDLEQPVNIELPIGF	104
Db	761	KEIQDKSEELHITSEKDKLFSEVHWKESRVQGLBEIGTKDKDLATQS-----	---	NY 812
Qy	105	EYLDYEY--	VRDVGVPDHDHTTYKVKNKNKNSLDKAVKHKSRLLI	IYENFPY---ILKRYV 159
Db	813	KSTDQEPNFKTLHMPD	EOYKQVLEBENRMNOEIVNLSKEA----	OKFOSLICALKTEL 868


```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7880
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7880

Query Match      6.4%; Score 132.5; DB 2; Length 1786;
Best Local Similarity 21.7%; Pred. No. 0.0057;
Matches 90; Conservative 63; Mismatches 143; Indels 119; Gaps 17;

QY 11 DKISLELYKYSI-----PFRNVIENVAEDCLNGLILESAAHNVSEVE--LARLKVQ 61
DB 998 DYLVSKQVKEVRIQBNRENSPHSSIEAIWEECKEIVKASSKSHOIELEQQIEKQAE 1057
QY 62 LKNAL-LNCIIISYRPHGIGYVLVKTDTLI-----DLEQPVNIELPIGEYLDYEVVDLG 116
DB 1058 VKGYKDENNRLKEKEHKQDQLLKEKETLIQOLKELOEKVLDVQIQHV-VEGKRLS 1116
QY 117 VDFDHT-YKVKNNKNSLDAVKTHKRLIYENPDFYILKRYVPCYTESFLDDIYLPFK 175
DB 1117 ELTQGVTCYKAKIKELETILETQKVERSHSAKLEQ-----DILEKES 1158
QY 176 IVEVTERIENHNFLFYKDESIVQLODALSATTLSALTQSNDRGSGILSSFLRQ-- 233
DB 1159 IILKLERNLK-----EFOEHQDSV-----KNTKDLNVKELKLEBIT 1196
QY 234 NSNNHKSIDSINRLN-----NDLSLQELARLKNLNEGMYTATPSASLEVIKYDLSYLK 289
DB 1197 QLTNNLQDMKHLQLKESSEETNRQETEKLEEL-----SASARTQNLKADLQKRE 1248
QY 290 EALALIKAKIGADTKPELTRSFNQAKGLNGDKGRSNYYDFLKGVEQVENACNLKLT 349
DB 1249 EDY-----ADLKEKLT-----DAKKQ-----IKQVQKEVS----- 1273
QY 350 KYFGLDMKFNLSLMLSEQKVERDIKLIYLSKYNOLIOSSSFDNEELAILKEKL 404
DB 1274 -----VMRDEDKLLR-IKINELEKKNQCSQELDMKQRTIQQLKEQL 1314

RESULT 8
US-09-248-796A-16380
; Sequence 16380, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16380
; LENGTH: 511
```

```
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-16380

Query Match      6.4%; Score 132; DB 2; Length 511;
Best Local Similarity 20.2%; Pred. No. 0.0009;
Matches 92; Conservative 72; Mismatches 149; Indels 142; Gaps 22;

QY 49 NVSEVELARLKVQLKNALLNCIIISYRPH-GIGYVLVK-TKDTLIDLEQPVNIELPIGEY 106
DB 41 DIHEKRMQR-NIEITKSIQITMYAKLNERVDYILEKLTSTELDSEKVM-----K 91
QY 107 LDYEVYRDLGVDFHITYKVKSNKNSLDAVKTHKSLRIIYENPDFYILKRYVPCYTESF 166
DB 92 LNSEFDPQEKEDYDKL-IKDKGLTLRKGLKDLKFDREI---EN-----TPCYNE-- 137
QY 167 LLDIYLFKEIYVEI-----ERRIENHNPLFYKDSLVQLODALSATTLSALTOS---- 217
DB 138 ----MIEDLFVQIKDDHPETKTGDGKLIYELKEHNRIDDLVLSKQTIKLDLQYKAQL 192
QY 218 --NDRGSGILSSFLRKQSNNSHKSIDS-----NL-----RNL 248
DB 193 IVSDDLHTGDFDRSFLNKDKPEEKENLDKPKAPRAKTTVTVTETINLPKPVENDTKEI 252
QY 249 NDLSL-----SOELARLKS-NLNNEGMYTATPSASLEVIK-----YD----- 284
DB 253 LDELEILPATKEPAKIPSDNLSKAAEFILKHPISCTEQOKDALIMTAFDLQLENKSDEAK 312
QY 285 -----LSYLKE-----ALALIKAKIGADTKPELTRSFNQAKGLNGD 322
DB 313 HIVHQSLLQVQQLSGNGKAAPANVINAIKLFESKIAAES-SPAKHAFLE----- 362
QY 323 KGDRSNYYDFLKG-----VQEOVENACNLK-----LTKYFGLDMKFNLSLMLSE 368
DB 363 --DVNQTFNHITKGEIIEKEQROANAANKETGEBEEBALIQLKALDDNTTELLVNIPOEG 420
QY 369 KVERDIKLIYLSKYNOLIOSSSFD--NEELAILK 401
DB 421 TKEYEIFTKLPFQNAIKTESIDENVKEPAKLK 455

RESULT 9
US-09-134-001C-4378
; Sequence 4378, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4378
; LENGTH: 878
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4378

Query Match      6.4%; Score 131.5; DB 2; Length 878;
Best Local Similarity 23.4%; Pred. No. 0.0023;
Matches 94; Conservative 79; Mismatches 144; Indels 85; Gaps 25;

QY 27 NVIEN-VAEDCLKNGLILESAAHNVSEVELARLKVQLKNALLNCI----- 70
DB 363 NNVSNTYTKDILPN---IEVYSQN-----EIIDLTNN-ENAKLINILNRLFKDDRNDKKE 414
QY 71 -ISYRFHGIGYVLVKTDTLIDLEQPVNIELPIGEYLDYEVYRDLGVDFDHTYKVKSN 129
```

Db 415 EIKTNLHNSKSLIKAKEDLENLQEKIN-QLPKLEKL--KHFNELG-----IGKLVQ 466

QY 130 NKNNSLDAVKIHKSLRIIYENFDYILKRYVPCYTESFLDIDYLFPEKIYVEIERRIENHNF 189

Db 467 GK-ISREEQYQNTQIILEDN-DISITNIIILFPNENYQIKHVE-IFDSIKNIIDNEN- 522

QY 190 LFYKDESILVOJODALSSATTSLSALTQSNDRGSGILSSFLRKQNSNNHSDISN-LRNL 248

Db 523 -----KLUKE-IKSMFTDLKOTTONEIEKIYVNW-----KEKKKNIKEINRAIKSL 568

QY 249 NDSLSQELARLKSNNLNGEMFYTAPS--ASLEVIKYDLSYLKEALALIKAKIGADTKEP 306

Db 569 DDI-----EGTKEDIAHE---YETOKQITSTIEPLETQLSRVKTSIETLENE-RIQLKED 620

QY 307 LTRSFNEQAKGLGNDGKGRSNIYDFLKGVOQVENACNLKLTIFYGLDMKFNSLIMLSE 366

Db 621 LKEIFEQKLNLCRVCKINRY-----LKKQV-----NIKQPYANVN---NLIEFLKE 667

QY 367 EOKV-ERDIKLIYLSKYN-----QLIOSSPFDNELAILKE 402

Db 668 ENGLGDSTLWKIKNHQSFPNFKFIKLIK----DRDSEAIYEE 705

RESULT 10

US-09-541-782-6

; Sequence 6, Application US/09541782

; Patent No. 6284480

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Berand, Christophe

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT APPLICATION NUMBER: US/09/541,782

; CURRENT FILING DATE: 2000-04-03

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 1073

; TYPE: PRT

; ORGANISM: S. pombe

US-09-541-782-6

Query Match 6.3%; Score 130.5; DB 2; Length 1073;

Best Local Similarity 19.5%; Pred. No. 0.004;

Matches 85; Conservative 68; Mismatches 167; Indels 115; Gaps 15;

QY 30 ENVAEDCLKNGLILESAAHNVSEVELARLKVQKNALLNCI--ISYRPHGIGYVLV--- 83

Db 320 ENIGRGAENKRAETGMINQSLTLGR-----VINALVEKAHHIYPRESKLTRLQDSLG 375

QY 84 -KTQDITLDLEQPVNIELPIGPEYLDYE-----YVRDLGVDFDHI 122

Db 376 GKTKTSMIVTVSSVTNLEETISTLEYAARAKSIRKPNQNLVPRKVLIKDLVDIERL 435

QY 123 TVYKSNNNKNSLDAVKIHKSLRIIYENFDYILKRYVPCYTESFLDIDYLFPEKIYVEIER 182

Db 436 KNDLNATRKNG-----VYLAESTYKELMD 460

QY 183 RIENHNFYKDESILVQLODA-LSSATTSLSALTQSNDRGSGILSSFLRKQNSNNHSD 241

Db 461 RVQNKDLLCQEQARKLEVLDLNVKSSREQLQVYKSNQEHKKEVEALQQLVNS---STE 517

QY 242 ISNLRNLNDSLSQELARLKSNNLNGEMFYTAPSASLEVIKYDLSYL---KEALALIKA 297

Db 518 LESVKSENEKLNELV-----LEIEKRKKYETNEAKITTVATDLSQYVRESKEYIASLYE 572

QY 298 KIG---ADTKE-----PLTRSENEQAKGLGNDGKGRSNIY---DPLKGV 336

Db 573 KLDRTERRNNKENNFWNLKFNLLTMLRSFHSGFT-----DETNGYFTLLDNFNASM 624

QY 337 QBOVENACN-----LKLTKYF-GLDMKFNSLIMLSEEQKVERDIKLIYLSKYNQLIQS 389

Db 390 SSFONEELAILKEKL 404

Db 685 LEHSLQDISMSQKL 699

RESULT 12

US-10-270-085-6

; Sequence 6, Application US/10270085

; Patent No. 6627408

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

Db 625 BELNTHSNQLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKDSKNSLLDA 684

QY 390 SSFONEELAILKEKL 404

Db 685 LEHSLQDISMSQKL 699

RESULT 11

US-09-723-820-6

; Sequence 6, Application US/09723820

; Patent No. 6468760

; GENERAL INFORMATION:

; APPLICANT: Nislow, Corey

; APPLICANT: Sakowicz, Roman

; APPLICANT: Berand, Christophe

; TITLE OF INVENTION: Antifungal Assay

; FILE REFERENCE: 1015

; CURRENT APPLICATION NUMBER: US/09/723,820

; CURRENT FILING DATE: 2000-11-28

; PRIOR APPLICATION NUMBER: 09/541,782

; PRIOR FILING DATE: 2000-04-03

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 1073

; TYPE: PRT

; ORGANISM: S. pombe

US-09-723-820-6

Query Match 6.3%; Score 130.5; DB 2; Length 1073;

Best Local Similarity 19.5%; Pred. No. 0.004;

Matches 85; Conservative 68; Mismatches 167; Indels 115; Gaps 15;

QY 30 ENVAEDCLKNGLILESAAHNVSEVELARLKVQKNALLNCI--ISYRPHGIGYVLV--- 83

Db 320 ENIGRGAENKRAETGMINQSLTLGR-----VINALVEKAHHIYPRESKLTRLQDSLG 375

QY 84 -KTQDITLDLEQPVNIELPIGPEYLDYE-----YVRDLGVDFDHI 122

Db 376 GKTKTSMIVTVSSVTNLEETISTLEYAARAKSIRKPNQNLVPRKVLIKDLVDIERL 435

QY 123 TVYKSNNNKNSLDAVKIHKSLRIIYENFDYILKRYVPCYTESFLDIDYLFPEKIYVEIER 182

Db 436 KNDLNATRKNG-----VYLAESTYKELMD 460

QY 183 RIENHNFYKDESILVQLODA-LSSATTSLSALTQSNDRGSGILSSFLRKQNSNNHSD 241

Db 461 RVQNKDLLCQEQARKLEVLDLNVKSSREQLQVYKSNQEHKKEVEALQQLVNS---STE 517

QY 242 ISNLRNLNDSLSQELARLKSNNLNGEMFYTAPSASLEVIKYDLSYL---KEALALIKA 297

Db 518 LESVKSENEKLNELV-----LEIEKRKKYETNEAKITTVATDLSQYVRESKEYIASLYE 572

QY 298 KIG---ADTKE-----PLTRSENEQAKGLGNDGKGRSNIY---DPLKGV 336

Db 573 KLDRTERRNNKENNFWNLKFNLLTMLRSFHSGFT-----DETNGYFTLLDNFNASM 624

QY 337 QBOVENACN-----LKLTKYF-GLDMKFNSLIMLSEEQKVERDIKLIYLSKYNQLIQS 389

Db 625 BELNTHSNQLLISMTKITEHFQSLDEALQASRSCAVPNSSLDLIVSELKDSKNSLLDA 684

QY 390 SSFONEELAILKEKL 404

Db 685 LEHSLQDISMSQKL 699

```

; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/10/270,085
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US/09/723,820
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: PRT
; ORGANISM: S.pombe
US-10-270-085-6

```

Query Match 6.3%; Score 130.5; DB 2; Length 1073;
 Best Local Similarity 19.5%; Pred. No. 0.004;
 Matches 85; Conservative 68; Mismatches 167; Indels 115; Gaps 15;

```

QY 30 ENVAEDCLKNGILLESAAHNVSEVELARLKVQLKALLNCI--ISYRPHGIGYVLV---- 83
Db 320 ENIGSGAENKARETGMINQSLLTLGR---VINALVEKAHHIPYRESKLTLLQDSLG 375

QY 84 -KTKDTLIDLEQPVNIELPIGPEYLDYE-----YVRDLGVDFDHI 122
Db 376 GKTKTSMIVTVSTNTNLEETISTLEYAARAKSIRKPNQNLVPRKVLKDLVDIERL 435

QY 123 TVYKSNNNKNSLDAVKIHKSLIYENFDYILKRVPCYTESFLDYLVPFKIYVEIER 182
Db 436 KNDLNATRKNKG-----VYLAESTYKELMD 460

QY 183 RIENHNFYKDESLOQLODA-LSSATTSLSALTQSNDRGSGILSSFLRKQNSNNHSD 241
Db 461 RVQNKDLLCOEQARKLEVLDAVYKSREQLQYVSKNSQHEKVEALQQLVNS---STE 517

QY 242 ISNLRNLNLSOELARLKSNNNEGMFYATPSPASLEVIKYDLSYL---KEALALIKA 297
Db 518 LESVKSENEKLNELV-----LEIEKRYKYETNEAKITVATDLSQYRESKEYIASLYE 572

QY 298 KIG---ADTKE-----PLTRSFNBQAKGLGNDGKGRSNYY-----DFLKG 336
Db 573 KLDRTERRNNKENENFNWLNKLLTWMRSFHGSFT-----DETNGYFTLLNDFNASM 624

QY 337 QOQVENACH-----LKLTKYP-GLDMKFNLSIMLSEQKVERDIKLIYKYNQLIQS 389
Db 625 EELLNTHSNQLLISMTKITTEHPQSLDEALQASRSCAVPNSSLDLIVSELKDSKNSLLDA 684

QY 390 SSFDNEELAILKEKL 404
Db 685 LEHSLODISMSQKL 699

```

```

RESULT 13
US-09-724-519-8
; Sequence 8, Application US/09724519
; Patent No. 6414121
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth
; APPLICANT: Finer, Jeffrey
; APPLICANT: Beraud, Christophe
; APPLICANT: Mak, John
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Methods of screening for modulators of
; TITLE OF INVENTION: cell proliferation and methods of diagnosing cell
; TITLE OF INVENTION: proliferation states
; FILE REFERENCE: 1014A
; CURRENT APPLICATION NUMBER: US/09/724,519
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 09/592,037
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/428,156

```

```

; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Human
US-09-724-519-8

```

Query Match 6.3%; Score 129.5; DB 2; Length 575;
 Best Local Similarity 22.4%; Pred. No. 0.0019;
 Matches 105; Conservative 63; Mismatches 175; Indels 125; Gaps 22;

```

QY 7 TKLIDKISSLELYKYSIP-FRNYIENVAEDCL-----KNGLIIL-----ESAHNVSEV- 53
Db 152 TEFSVKVSLLEIYNEELFDLLNPSDSVSRQLQMPDPRNKRGVIIKGLLEIIVHNKDEVY 211

QY 54 ---ELARLKVQLKALLNCIISYRPHGIGYVLVTKOTFLIDLEQPVNIELPIGPEYLDYE 110
Db 212 GILEKGAARKTAAATLMAA-YSSRSHSVFVTIHKMETTIDGEEELVKIG---KLNLDVLA 267

QY 111 YVRDLG-----VDFDHIITYKVKSNKNSLDAVKIHKSLIYENFDYILKR--YVPCYTE 164
Db 268 GSENIGRSGAYD-----KRAREAGNINQSLLTLGRVITALVERTPHVP-VRE 313

QY 165 SFLLDIYLFKEIYVEIERRIENHNFLFYKDESLOQLODALSSAT-TLSALTQSNDRGS 223
Db 314 SKLIRI-----LQDSLGRFTSTIATISP-----AS 340

QY 224 GILSFLRKQNSNNHSDISNLRNLNDSL-----SOELARLKSML-----NNEGMPY 270
Db 341 LNLEETLSTLEYAHRANKILNKPEVNQKLTKKALIKETIEIERLKRDLAAAREKNGVYI 400

QY 271 TATSPASLEVIKYDLSYLKEALALIKAGI-----DTKEPL----- 307
Db 401 S---EENFRVMSGKLTVOEQIIVELIEKIGAVEBELNRVTELFMNKNELDQCKSDLQNK 457

QY 308 TRSFNBQAKGLGNDGKGRSNYY--DFLKGQEQOVENACNLKL-----TK-YFLGDMKF 358
Db 458 TQLETTQKHGTEKGLVKEEYTSALSTEEKLDHAASKLLNVTETTKDVSGLHSL 517

QY 359 NSLIMLSEQKVERDI---KLIELYSKYNQLIQSSSFDNEELAILKEK 403
Db 518 DRKRAVDQHNAEQDIFGKNLNSLFNNMEELIKGSLQTEGKLISEE 565

```

```

RESULT 14
US-09-592-037-8
; Sequence 8, Application US/09592037
; Patent No. 6437115
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth
; APPLICANT: Finer, Jeffrey
; APPLICANT: Beraud, Christophe
; APPLICANT: Mak, John
; APPLICANT: Sakowicz, Roman
; TITLE OF INVENTION: Methods of screening for modulators of
; TITLE OF INVENTION: cell proliferation and methods of diagnosing cell
; TITLE OF INVENTION: proliferation states
; FILE REFERENCE: 1014A
; CURRENT APPLICATION NUMBER: US/09/592,037
; CURRENT FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/428,156
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Human
US-09-592-037-8

```

Query Match 6.3%; Score 129.5; DB 2; Length 575;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 20.3361 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-6
Perfect score: 470
Sequence: 1 LQGFVPSYVPISSISLKG.....SIKNEWLLVPLKPHWAFS 91

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*
9: Geneseqp2005s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	91	8	ADU98747 Borrelia
2	468	99.6	275	2	AAV19955 B. burgdo
3	468	99.6	290	2	AAV19954 B. burgdo
4	468	99.6	290	8	ADU98749 Borrelia
5	193	41.1	384	7	ABO72983 Pseudomon
6	186.5	39.7	573	5	ABBS4794 Lactococc
7	179	38.1	309	9	ABM96406 M. xanthu
8	161.5	34.4	577	7	ADC95590 E. faeciu
9	158.5	33.7	569	4	AAU34910 Enterococ
10	158.5	33.7	571	7	ADH87483 Streptoco
11	157	33.4	575	5	ABF27731 Streptoco
12	153	32.6	273	6	ADB07854 Alloioococ
13	153	32.6	334	6	ADB07856 Alloioococ
14	153	32.6	352	6	ADB07858 Alloioococ
15	152.5	32.4	320	6	ADB07852 Alloioococ
16	152.5	32.4	575	5	ABP27730 Streptoco
17	152.5	32.4	575	5	ABP29826 Streptoco
18	152.5	32.4	575	8	ADV88795 Streptoco
19	152.5	32.4	575	8	ADV80048 Streptoco
20	152.5	32.4	575	8	ADV82185 Streptoco
21	147	31.3	349	4	AAU46688 Propionib
22	147	31.3	349	6	ABM43207 Propionib
23	138	29.4	312	4	AAU33717 Pseudomon
24	138	29.4	312	6	ABU15624 Protein e

25	138	29.4	318	7	ABO68427	AbO68427 Pseudomon
26	136	28.9	417	4	AAU40421	Aau40421 Propionib
27	136	28.9	417	6	ABM36940	Abm36940 Propionib
28	135.5	28.8	296	8	ADG32088	Adg32088 Mutant B
29	130	27.7	302	6	ADB11550	Adb11550 Alloioococ
30	127	27.0	356	7	ABO68411	AbO68411 Pseudomon
31	127	27.0	380	6	ABU21819	Abu21819 Protein e
32	126	26.8	427	7	ABO73023	AbO73023 Pseudomon
33	124.5	26.5	300	5	ABB48738	Abb48738 Listeria
34	124	26.4	303	6	ADB07978	AdB07978 Alloioococ
35	124	26.4	326	6	ADB07980	AdB07980 Alloioococ
36	108	23.0	314	6	ABU39563	Abu39563 Protein e
37	88	18.7	317	7	ABO82583	AbO82583 Pseudomon
38	87	18.5	333	6	ABM69068	Abm69068 Photorhab
39	86	18.3	338	7	ABO63655	AbO63655 Klebsiell
40	76.5	16.3	903	6	ABU49526	Abu49526 Protein e
41	76	16.2	361	7	ABO84191	AbO84191 Pseudomon
42	74	15.7	319	5	ABP60851	Abp60851 Buchnera
43	74	15.7	573	5	ABP39811	Abp39811 Staphyloc
44	74	15.7	573	8	ADS05795	Ads05795 Staphyloc
45	73.5	15.6	687	5	ABB92975	Abb92975 Herbicida

ALIGNMENTS

RESULT 1
ADU98747
ID ADU98747 standard; protein; 91 AA.
XX
AC ADU98747;
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide segid 6.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
FN WO2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
XX WPI; 2004-834155/82.
DR N-PSDB; ADU98746.
PT Immunizing a subject against Borrelia burgdorferi infection comprises
PT providing to the subject at least one Borrelia antigen or its fragment.
PS Claim 27; SEQ ID NO 6; 121pp; English.
CC The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide
CC encoding a polypeptide for an ability to produce an immune response;
CC preparing a vaccine; vaccinating a subject; treating a subject infected

XX
DR WPI: 2001-611495/70.
DR N-PSDB; AAS52769.

XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.

XX
XX Example 3; SEQ ID NO 10503; Silpp; English.

XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC [ftp.wipo.int/pub/published pct](http://ftp.wipo.int/pub/published/pct) sequences

are not form of the granted specification; one was examined in electronic format directly from WIPO at ftp.wipo.int/pub/published pct sequences

Quest Match	33.7%	Pred. No. 2.2e-10;	
Best Local Similarity	37.1%	Pred. No. 2.2e-10;	
Matches	33;	Conservative	15; Mismatches 38;
			Indels 3; Gaps 1

D0 402 GIVPVSNDVDSLEDLSOAGK---KITGLEPGAGWAAAEKTKEAYPNLKOWSVETSSS 458

Db 459 GANTVALQAIKNNEDIVITGSPHWMFA 487

ADH87483
ID ADH87483 standard; protein; 571 AA.

Enterococcus faecalis infection; transcription regulatory element; antibacterial.

00 MICELLOCCUS LACCAIS.
00 XX
00 PN US6617156-B1.

PF 13-AUG-1998; 98US-00134000.

PR 15-AUG-1997; 9TUS-0055778P.
XX
XX
XX

XX Doucette-Stamm LA, Bush D;
PI
vv

DR N-PSDB; ADH84078.

XX New nucleic acid comprising a sequence encoding an Enterococcus faecalis
PT polypeptide, useful for preparing a composition for diagnosing or
PT treating E. faecalis infection.
XX
PS Disclosure; SEQ ID NO 5368; 193pp; English.
XX
CC The invention relates to Enterococcus faecalis polynucleotides and
CC polypeptides. The invention also relates to a recombinant expression
CC vector comprising a polynucleotide operably linked to a transcription
CC regulatory element, a cell comprising a recombinant vector, a method for
CC producing an E. faecalis polypeptide, an isolated nucleic acid comprising
CC a sequence not given in the specification, a recombinant vector
CC comprising the nucleic acid and a cell comprising the recombinant vector.
CC The polynucleotides can be used to detect the presence of E. faecalis in
CC a sample. The sequences are useful for preparing a composition for
CC diagnosing or treating Enterococcus faecalis infection. This sequence
CC represents an E. faecalis polypeptide of the invention.
XX
SQ Sequence 571 AA;

Query Match 33.7%; Score 158.5; DB 7; Length 571;
Best Local Similarity 37.1%; Pred. No. 2.2e-10;
Matches 33; Conservative 15; Mismatches 38; Indels 3; Gaps 1;
QY 3 GFVPSYVPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62
DB 404 GIVPSYMDVSDIEDLSDAQK---KITGIEFGVVAANAETKAYPNLKDMSVETSSS 460
QY 63 SYMLASLSSIKRNEWILVPLWKPWFAS 91
DB 461 GANTVALGQAIRKNEEDIVTGWSPHMF 489

RESULT 11
ABP27731
ID ABP27731 standard; protein; 575 AA.
XX
AC ABP27731;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 4638.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus pyogenes.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI; 2002-352536/38.
DR N-PSDB; ABN68362.
XX
PT New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX

PS Claim 1; Page 3626-3627; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
SQ Sequence 575 AA;

Query Match 33.4%; Score 157; DB 5; Length 575;
Best Local Similarity 36.7%; Pred. No. 3.5e-10;
Matches 33; Conservative 17; Mismatches 36; Indels 4; Gaps 2;
QY 3 GFVPSYVPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 61
DB 407 GLAVPKYMTVDNSIEDLSKQADQ---KITGIEFGAGIMAAQKTUKEYHNLSWELVAAS 463
QY 62 ESMVLASLSSIKRNEWILVPLWKPWFAS 91
DB 464 TGAMTTSILDQAIRKNDPIVVTANSPHMF 493

RESULT 12
ADB07854
ID ADB07854 standard; protein; 273 AA.
XX
AC ADB07854;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloicoccus otitis antigenic protein SEQ ID NO:1794.
XX
KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.
XX
OS Alloicoccus otitis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
DR WPI; 2003-505284/47.
DR N-PSDB; ADB07853.
XX
PT New Alloicoccus otitis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
PS Claim 33; SEQ ID NO 1794; 1019pp; English.
XX
CC The present invention describes an isolated polynucleotide (I) of

Alloccoccus otitidis genomic DNA, which encodes an antigenic protein. Alloccoccus otitidis is a Gram-positive bacterium. Also described: (1) an isolated polypeptide that is encoded by the polynucleotide (1); (2) an expression vector comprising the novel isolated polynucleotide (1), its complement, degenerate variant or fragment; (3) a genetically engineered host cell, transfected, transformed or infected with the vector of (2); (4) an antibody specific for the polypeptide of (1); (5) an immunogenic composition comprising the polypeptide, its complement, biological equivalent or fragment, or the polynucleotide that is comprised in the expression vector; (6) a pharmaceutical composition comprising the polypeptide of (1) and a carrier; (7) a protein chip comprising an array of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against Alloccoccus otitidis by administering to a host the immunogenic composition; (9) detecting and/or identifying Alloccoccus otitidis in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (1) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting Alloccoccus otitidis. The present sequence represents an Alloccoccus otitidis antigen protein from the present invention.

```

XX
SQ      Sequence 273 AA;

Query Match      32.6%; Score 153; DB 6; Length 273;
Best Local Similarity 39.1%; Pred. No. 4e-10;
Matches 36; Conservative 19; Mismatches 31; Indels 6; Gaps 4

QY      1  LOGFVPSYVP-ISSIGELKGGDKFNK-MIGIDAGAGTQIVTFOALNYYGLSKEYELV 58
         44  LIGLVPEYMEGLSDIEDLE---DQIEDKTVGIERGAGTVEAAREALDYG-KEDWEVI 99

Db

QY      59  PSSESVMLASLDSSIKENWILVPLWKDHPAF 90
         ||| : : : |||
Db      100  ASSGGMVTELQAINNEEDIIVGTWTPHWRP 131
         ||| : : : |||

RESULT 13
ADB07856
ID      ADB07856 standard; protein; 334 AA.

```

Claim 33; SEQ ID NO 1796; 1019pp; English.

The present invention describes an isolated polynucleotide (I) of *Alloccoccus otitidis* genomic DNA, which encodes an antigenic protein. *Alloccoccus otitidis* is a Gram-positive bacterium. Also described: (1) an isolated polypeptide that is encoded by the polynucleotide (I); (2) an expression vector comprising the novel isolated polynucleotide (I), its complement, degenerate variant or fragment; (3) a genetically engineered host cell, transfected, transformed or infected with the vector of (2); (4) an antibody specific for the polypeptide of (1); (5) an immunogenic composition comprising the polypeptide, its complement, biological equivalent or fragment, or the polynucleotide that is comprised in the expression vector; (6) a pharmaceutical composition comprising the polypeptide of (1) and a carrier; (7) a protein chip comprising an array of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against *Alloccoccus otitidis* by administering to a host the immunogenic composition; (9) detecting and/or identifying *Alloccoccus otitidis* in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (1) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting *Alloccoccus otitidis*. The present sequence represents an *Alloccoccus otitidis* antigen protein from the present invention.

RESULT 14
ADB07858
ID ADB07858 standard; protein; 352 AA.
XX
XX ADB07858;
XX
DT 20-NOV-2003 (first entry)
XX
XX DE Alloococcus otitis antigenic protein SEQ ID NO:1798.
DE
XX DE Alloococcus otitidis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection.

```
PI Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
DR N-PSDB; ADB07857.
XX
XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
XX Claim 33; SEQ ID NO 1798; 1019pp; English.
XX
XX The present invention describes an isolated polynucleotide (1) of
CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.
CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiooccus otitidis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
CC otitidis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloiooccus
CC otitidis. The present sequence represents an Alloiooccus otitidis
CC antigen protein from the present invention.
XX
XX Sequence 352 AA;
XX
XX Query Match 32.6%; Score 153; DB 6; Length 352;
XX Best Local Similarity 39.1%; Pred. No. 5.6e-10;
XX Matches 36; Conservative 19; Mismatches 31; Indels 6; Gaps 4;
XX
XX QY 1 LOGFVVPSTVP-SSISLKGKDKFNK-MIGIDAGAGTQIVTQALNYYGLSKEYELV 58
XX 123 LGLVVPYEWGDSIEDLE---DQEDKTVIGERGAGTVEAREALEDDYG-KEDWEVI 178
XX
XX QY 59 PSSESVMLASLSSIKRNEWILVPLWKPHWAF 90
XX 179 ASSSGAMVTELQAANNEEDIIVTGWTPHKF 210
XX
XX Db
XX
XX RESULT 15
XX ADB07852
XX ID ADB07852 standard; protein; 320 AA.
XX AC ADB07852;
XX
XX XX 20-NOV-2003 (first entry)
XX
XX DE Alloiooccus otitis antigenic protein SEQ ID NO:1792.
XX
XX KW Alloiooccus otitidis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection.
XX
XX OS Alloiooccus otitis.
XX
XX PN WO2003048304-A2.
XX
XX PD 12-JUN-2003.
XX
XX PF 25-NOV-2002; 2002WO-US036123.
XX
XX 29-NOV-2001; 2001US-0333777P.
XX 18-NOV-2002; 2002US-0426742P.
XX
XX (AMHP ) WYETH HOLDINGS CORP.
XX
XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX
XX WPI; 2003-505284/47.
XX N-PSDB; ADB07851.
XX
XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
XX Claim 33; SEQ ID NO 1792; 1019pp; English.
XX
XX The present invention describes an isolated polynucleotide (1) of
CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.
CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiooccus otitidis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiooccus
CC otitidis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloiooccus
CC otitidis. The present sequence represents an Alloiooccus otitidis
CC antigen protein from the present invention.
XX
XX Sequence 320 AA;
XX
XX Query Match 32.4%; Score 152.5; DB 6; Length 320;
XX Best Local Similarity 36.4%; Pred. No. 5.7e-10;
XX Matches 32; Conservative 17; Mismatches 36; Indels 3; Gaps 1;
XX
XX QY 3 GFVVPSTVPISISLKGKDKFNKMGIDAGAGTQIVTQALNYYGLSKEYELVPSSE 62
XX 122 GLGVPTMEVDSIADLT---DEAQTTIGERAGVGVSQAQLEDDYDNLSDWTATSSS 178
XX
XX Db
XX
XX QY 63 SVMLASLSSIKRNEWILVPLWKPHWAF 90
XX 179 GAMVTELETAANNEEDIVVTAWSPHWKF 206
XX
XX Db
XX
XX Search completed: January 24, 2006, 19:44:28
XX Job time : 22.3361 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 5.60996 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-6
Perfect score: 470
Sequence: 1 LQGFVVPVPSISSSELKKG.....SIKRNWILVPLWKPHWAFS 91

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/1aa/5 COMB.pap:*
- 2: /cgn2_6/ptodata/1/1aa/6 COMB.pap:*
- 3: /cgn2_6/ptodata/1/1aa/H COMB.pap:*
- 4: /cgn2_6/ptodata/1/1aa/PTUS COMB.pap:*
- 5: /cgn2_6/ptodata/1/1aa/RE COMB.pap:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	99.6	275	2	US-09-830-230A-362
2	468	99.6	290	2	US-09-830-230A-361
3	193	41.1	384	2	US-09-252-991A-21729
4	179	38.1	309	2	US-09-902-540-15605
5	161.5	34.4	577	2	US-09-107-532A-5217
6	158.5	33.7	571	2	US-09-134-000C-5368
7	138	29.4	318	2	US-09-252-991A-17173
8	127	27.0	356	2	US-09-252-991A-17157
9	126	26.8	427	2	US-09-252-991A-21769
10	88	18.7	317	2	US-09-252-991A-31329
11	86	18.3	338	2	US-09-489-039A-10172
12	76	16.2	361	2	US-09-252-991A-32937
13	74	15.7	573	2	US-09-134-001C-4656
14	69	14.7	426	2	US-09-807-258-26
15	68	14.5	317	2	US-09-489-039A-9963
16	67.5	14.4	1208	2	US-09-540-236-3604
17	67.5	14.4	1218	2	US-09-198-452A-98
18	67.5	14.4	1410	2	US-09-438-185A-84
19	67	14.3	289	2	US-09-807-258-22
20	67	14.3	379	2	US-09-248-796A-18875
21	66	14.0	251	2	US-09-107-532A-5242
22	66	14.0	617	2	US-09-565-264-2
23	66	14.0	617	2	US-09-565-264-4
24	66	14.0	646	2	US-09-252-991A-16819
25	66	14.0	1059	2	US-09-134-000C-5600
26	65.5	13.9	364	2	US-08-818-112-73
27	65.5	13.9	364	2	US-08-818-111-74

28	65.5	13.9	364	2	US-09-056-556-73	Sequence 73, Appl
29	65.5	13.9	364	2	US-09-072-596-74	Sequence 74, Appl
30	65.5	13.9	364	2	US-09-072-967-73	Sequence 73, Appl
31	65.5	13.9	364	2	US-10-193-002-74	Sequence 74, Appl
32	65.5	13.9	364	2	US-10-084-843-73	Sequence 73, Appl
33	65	13.8	1215	2	US-09-134-001C-5319	Sequence 5319, Ap
34	64	13.6	442	2	US-08-821-994-66	Sequence 66, Appl
35	63.5	13.5	837	2	US-09-122-126B-2	Sequence 2, Appl
36	63.5	13.5	837	2	US-09-634-286A-2	Sequence 2, Appl
37	63.5	13.5	837	2	US-10-247-685-2	Sequence 2, Appl
38	63.5	13.5	837	2	US-09-949-002-380	Sequence 380, App
39	63.5	13.5	837	2	US-10-012-231A-317	Sequence 317, App
40	63.5	13.5	837	2	US-10-015-389A-317	Sequence 317, App
41	63.5	13.5	837	2	US-10-006-768A-317	Sequence 317, App
42	63.5	13.5	837	2	US-10-015-671A-317	Sequence 317, App
43	63.5	13.5	837	2	US-10-015-393A-317	Sequence 317, App
44	63.5	13.5	837	2	US-10-011-833A-317	Sequence 317, App
45	63.5	13.5	837	2	US-10-006-041A-317	Sequence 317, App

ALIGNMENTS

RESULT 1

US-09-830-230A-362
; Sequence 362, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 362
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-362

Query Match 99.6%; Score 468; DB 2; Length 275;
Best Local Similarity 98.9%; Pred. No. 3e-52;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQGFVVPVPSISSSELKKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 60
DB 94 IQGFVVPVPSISSSELKKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 153
QY 61 SESVMLASLDSISKKNWILVPLWKPHWAFS 91
DB 154 SESVMLASLDSISKKNWILVPLWKPHWAFS 184

RESULT 2

US-09-830-230A-361
; Sequence 361, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A

; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 361
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-361

Query Match 99.6%; Score 468; DB 2; Length 290;
Best Local Similarity 98.9%; Pred. No. 3.2e-52;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQGFVVPSPYPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYYGLSKYELVPS 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 109 LQGFVVPSPYPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYYGLSKYELVPS 168
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 SESVMLASLDSSIKRNEWILVPLWKPHWAPS 91
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 169 SESVMLASLDSSIKRNEWILVPLWKPHWAPS 199
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 3
US-09-252-991A-21729
; Sequence 21729, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21729
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21729

Query Match 41.1%; Score 193; DB 2; Length 384;
Best Local Similarity 42.7%; Pred. No. 2.1e-16;
Matches 38; Conservative 20; Mismatches 29; Indels 2; Gaps 1;

QY 3 GFVVPSPYPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYYGLSKYELVPS 62
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 215 GLIVPEYVANSIADLQAQDAFGGRVVGIDAGAGVMIKTDAIKQYGL--DYKLVASGG 272
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAPS 91
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 273 SGMIAEIATRAENEKPPVVTGIPHWMPA 301
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 4
US-09-902-540-15605
; Sequence 15605, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 15605
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-15605

Query Match 38.1%; Score 179; DB 2; Length 309;
Best Local Similarity 41.6%; Pred. No. 1e-14;
Matches 37; Conservative 16; Mismatches 34; Indels 2; Gaps 1;

QY 3 GFVVPSPYPISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYYGLSKYELVPS 62
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 138 GLVVPADLDTISIQNGKAKDLNQSIVGIDSGAGIMTTTEKATAYKL--DLKLVPSGG 195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAPS 91
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 196 PAMTAEIKDAIAKRPVVVTGMPKPKFA 224
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 5
US-09-107-532A-5217
; Sequence 5217, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:

```
; ORGANISM: Enterococcus faecium
;
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...577
; SEQUENCE DESCRIPTION: SEQ ID NO: 5217:
US-09-107-532A-5217

Query Match      34.4%; Score 161.5; DB 2; Length 577;
Best Local Similarity 37.1%; Pred. No. 4.5e-12;
Matches 33; Conservative 18; Mismatches 35; Indels 3; Gaps 1;

QY      3 GFVPSYVPISSELKGGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      410 GLAPYMDVNSIDELTQAGK---KIIEPGAGVVTAAENTIOKYNLKDWSVETSSS 466
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

QY      63 SVMLASLSDSSIKRNEWILVPLWKPHWAFS 91
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      467 GANTVALGQAIKKHEPIVVTGTPHMFPA 495
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

RESULT 6
US-09-134-000C-5368
; Sequence 5368, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5368
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5368

Query Match      33.7%; Score 158.5; DB 2; Length 571;
Best Local Similarity 37.1%; Pred. No. 1.1e-11;
Matches 33; Conservative 15; Mismatches 38; Indels 3; Gaps 1;

QY      3 GFVPSYVPISSELKGGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      404 GIVPSYMDVDSIEDLSQAGK---KITGIEPGAGVAAAEKTKAEPNLDKWSVETSSS 460
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

QY      63 SVMLASLSDSSIKRNEWILVPLWKPHWAFS 91
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      461 GANTVALGQAIKNNEDIVITGWSPHMFPA 489
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

RESULT 7
US-09-252-991A-17173
; Sequence 17173, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17173
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17173
```

```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17173

Query Match      29.4%; Score 138; DB 2; Length 318;
Best Local Similarity 33.7%; Pred. No. 2.2e-09;
Matches 30; Conservative 24; Mismatches 29; Indels 6; Gaps 3;

QY      4 FVPSYVP---ISSISELKGGDKFKNMIGIDAGAGTQIVTEQAL--NYGLSKEYELV 58
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      123 YAVPSYVAEGKLPADIARFKDKLGGRIYAIPEPGSGSNRITRKWIDNRFGL-KGFQLV 181
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

QY      59 PSSESVMLASLSDSSIKRNEWILVPLWKPH 87
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      182 ESSRAGMLTAVKRAIKRQWVFFGWKPH 210
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

RESULT 8
US-09-252-991A-17157
; Sequence 17157, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17157
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17157

Query Match      27.0%; Score 127; DB 2; Length 356;
Best Local Similarity 34.5%; Pred. No. 7e-08;
Matches 30; Conservative 20; Mismatches 31; Indels 6; Gaps 3;

QY      6 VPSYV---PISSISELKGGDKFKNMIGIDAGAGTQIVTEQAL--NYGLSKEYELVPS 60
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      161 VPQYVDGGLSPADIARFKDKLGNKIYGIIEPGDGNRVAQSMIDKNFELGK-FKLVES 219
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

QY      61 SESVMLASLSDSSIKRNEWILVPLWKPH 87
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
Db      220 SEAGMLSQVQRAIRRNQWVFLGWEPH 246
      :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:

RESULT 9
US-09-252-991A-21769
; Sequence 21769, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21769
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21769
```


Search completed: January 24, 2006, 20:02:39
Job time : 5.60996 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 15.9338 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-6

Perfect score: 470

Sequence: 1 LQGFVPSYVPISSISLKG.....SIKRNWILVPLMKPHWAFS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries .

Database :

- Published Applications AA Main:*
- 1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	100.0	91	5	US-10-688-058-6
2	468	99.6	275	5	US-10-994-726-362
3	468	99.6	290	5	US-10-688-058-8
4	468	99.6	290	5	US-10-994-726-361
5	192	40.9	871	4	US-10-156-761-14244
6	158.5	33.7	569	3	US-09-815-242-10503
7	157	33.4	575	5	US-10-474-792-100
8	153	32.6	273	5	US-10-501-282-1794
9	153	32.6	334	5	US-10-501-282-1796
10	153	32.6	352	5	US-10-501-282-1798
11	152.5	32.4	320	5	US-10-501-282-1792
12	138	29.4	312	3	US-09-815-242-5213
13	138	29.4	312	4	US-10-282-122A-43548
14	130	27.7	302	5	US-10-501-282-5510
15	127	27.0	380	4	US-10-282-122A-49743
16	124	26.4	303	5	US-10-501-282-1918
17	124	26.4	326	5	US-10-501-282-1920
18	108	23.0	314	4	US-10-282-122A-67487
19	87	18.5	332	4	US-10-156-761-10018
20	76.5	16.3	903	4	US-10-282-122A-77450
21	74	15.7	319	4	US-10-032-201B-200
22	74	15.7	773	4	US-10-724-972A-5090
23	71.5	15.2	773	4	US-10-437-963-162371
24	70	14.9	319	4	US-10-032-201B-201
25	69.5	14.8	581	4	US-10-437-963-122382
26	69.5	14.8	839	3	US-09-912-788-2
27	69.5	14.8	1178	4	US-10-282-122A-52187

28	69	14.7	1161	4	US-10-282-122A-52986	Sequence 52986, A
29	68.5	14.6	869	5	US-10-732-923-6769	Sequence 6769, Ap
30	68.5	14.6	1106	5	US-10-501-282-466	Sequence 466, App
31	68.5	14.6	1128	5	US-10-501-282-468	Sequence 468, App
32	68.5	14.6	1192	5	US-10-501-282-470	Sequence 470, App
33	68.5	14.6	1219	5	US-10-501-282-472	Sequence 472, App
34	68	14.5	299	4	US-10-282-122A-60579	Sequence 60579, A
35	68	14.5	424	4	US-10-282-122A-48661	Sequence 48661, A
36	68	14.5	461	5	US-10-482-076-6	Sequence 271, App
37	68	14.5	461	5	US-10-482-076-271	Sequence 51577, A
38	67.5	14.4	1182	4	US-10-282-122A-51577	Sequence 51577, A
39	67.5	14.4	1203	4	US-10-282-122A-45530	Sequence 45530, A
40	67.5	14.4	1218	4	US-10-289-762-98	Sequence 98, Appl
41	67.5	14.4	1396	4	US-10-282-122A-55213	Sequence 55213, A
42	67.5	14.4	1397	4	US-10-282-122A-54874	Sequence 54874, A
43	67.5	14.4	2560	4	US-10-276-774-1774	Sequence 1774, Ap
44	67.5	14.4	2923	3	US-09-788-711A-4	Sequence 4, Appli
45	67.5	14.4	2923	3	US-09-916-849A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-10-688-058-6
; Sequence 6, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:00305
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-6

Query Match 100.0%; Score 470; DB 5; Length 91;
Best Local Similarity 100.0%; Pred. No. 3e-48;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	LQGFVPSYVPISSISLKGDKFKNKWIGIDAGAGTQIVTEQALNYGLSKVELVPS	60
Db	1	LQGFVPSYVPISSISLKGDKFKNKWIGIDAGAGTQIVTEQALNYGLSKVELVPS	60
Qy	61	SESVMASLSDSIKRNEMILVPLMKPHWAFS	91
Db	61	SESVMASLSDSIKRNEMILVPLMKPHWAFS	91

RESULT 2
US-10-994-726-362
; Sequence 362, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483

```
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 362
; LENGTH: 275
; TYPE: PRN
; ORGANISM: Borrelia burgdorferi
US-10-994-726-362

Query Match          99.6%; Score 468; DB 5; Length 275;
Best Local Similarity 98.9%; Pred. No. 2.1e-47;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 60
Db 94 IQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 153

QY 61 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 91
Db 154 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 184

RESULT 3
US-10-688-058-8
; Sequence 8, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 290
; TYPE: PRN
; ORGANISM: Borrelia burgdorferi
US-10-688-058-8

Query Match          99.6%; Score 468; DB 5; Length 290;
Best Local Similarity 98.9%; Pred. No. 2.3e-47;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 60
Db 109 IQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 168

QY 61 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 91
Db 169 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 199

RESULT 4
US-10-994-726-361
; Sequence 361, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
```

```
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 361
; LENGTH: 290
; TYPE: PRN
; ORGANISM: Borrelia burgdorferi
US-10-994-726-361

Query Match          99.6%; Score 468; DB 5; Length 290;
Best Local Similarity 98.9%; Pred. No. 2.3e-47;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 60
Db 109 IQGFVVPSPVPISSISELKGDKFKNKMIGIDAGAGTQIVTEQALNYYGLSKEYELVPS 168

QY 61 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 91
Db 169 SESVMLASLDSSIKRNEWILVPLWKPHWAFS 199

RESULT 5
US-10-156-761-14244
; Sequence 14244, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14244
; LENGTH: 871
; TYPE: PRN
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14244

Query Match          40.9%; Score 192; DB 4; Length 871;
Best Local Similarity 45.5%; Pred. No. 1.2e-13;
Matches 40; Conservative 15; Mismatches 31; Indels 2; Gaps 2;

QY 6 VPSYV-PISSISELKGDKFKNKMIGIDAGAG-TQIVTEQALNYYGLSKEYELVPSSES 63
Db 699 VPSYVKDVKSLADLKGKGTEDGKIIEPGTGEMNLLKTKVLPGYGLDKEYKVVDGSGTP 758

QY 64 VMLASLDSSIKRNEWILVPLWKPHWAFS 91
Db 759 AMLAEKRAYAKKEFVAVVLMSPHWAFS 786
```

RESULT 6

US-09-815-242-10503
 ; Sequence 10503, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 10503
 ; LENGTH: 569
 ; TYPE: PRT
 ; ORGANISM: Enterococcus faecalis
 US-09-815-242-10503

Query Match 33.7%; Score 158.5; DB 3; Length 569;
 Best Local Similarity 37.1%; Pred. No. 6.9e-10;
 Matches 33; Conservative 15; Mismatches 38; Indels 3; Gaps 1;

QY 3 GFWVPSYVPISSISLKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62

DB 402 GIVPSYMDVDSIEDLSDAQK---KITGIEPGAGVAAAEKTKAAYPNLKDWSVETSSS 458

QY 63 SVMLASLSSIKRNEWILVPLWKPWF 91

DB 459 GANTVALGQAIKNNEDIVTGWSPHMFA 487

RESULT 7

US-10-474-792-100
 ; Sequence 100, Application US/10474792
 ; Publication No. US20040236072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Olmsted, Stephen
 ; APPLICANT: Zagursky, Robert
 ; APPLICANT: Nickbarg, Elliot
 ; APPLICANT: Winter, Lourie
 ; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES

FILE REFERENCE: AM 100399
 ; CURRENT APPLICATION NUMBER: US/10/474,792
 ; CURRENT FILING DATE: 2003-10-14
 ; NUMBER OF SEQ ID NOS: 674
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 100
 ; LENGTH: 575
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pyogenes
 US-10-474-792-100

Query Match 33.4%; Score 157; DB 5; Length 575;
 Best Local Similarity 36.7%; Pred. No. 1.1e-09;
 Matches 33; Conservative 17; Mismatches 36; Indels 4; Gaps 2;

QY 3 GFWVPSYVPISSISLKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 61

DB 407 GLAVPKYMTDVNSIEDLSKQADQ---KITGIEPGAGVAAAKTKLKEYHNLSSWELVAAS 463

QY 62 ESMVLASLSSIKRNEWILVPLWKPWF 91

DB 464 TGAMTTSILDQAKKKDPVVVTAWSPHMFA 493

RESULT 8

US-10-501-282-1794
 ; Sequence 1794, Application US/10501282
 ; Publication No. US20050203280A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MCMICHAEL, JOHN CALHOUN
 ; APPLICANT: ZAGURSKY, ROBERT JOHN
 ; APPLICANT: RUSSELL, DAVID PARRISH
 ; APPLICANT: FLETCHER, LEAH DIANE
 ; TITLE OF INVENTION: ALLOTOCOCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
 ; FILE REFERENCE: AM100780 L2
 ; CURRENT APPLICATION NUMBER: US/10/501,282
 ; CURRENT FILING DATE: 2004-07-09
 ; PRIOR APPLICATION NUMBER: 60/333,777
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: 60/426,742
 ; PRIOR FILING DATE: 2002-11-18
 ; PRIOR APPLICATION NUMBER: PCT/US02/36123
 ; PRIOR FILING DATE: 2002-11-25
 ; NUMBER OF SEQ ID NOS: 6653
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 1794
 ; LENGTH: 273
 ; TYPE: PRT
 ; ORGANISM: Alloiococcus otitidis
 US-10-501-282-1794

Query Match 32.6%; Score 153; DB 5; Length 273;
 Best Local Similarity 39.1%; Pred. No. 1.2e-09;
 Matches 36; Conservative 19; Mismatches 31; Indels 6; Gaps 4;

QY 1 LOGFVVPVVP-ISSISLKGKDKFKNK-MIGIDAGAGTQIVTEQALNYGLSKEYELV 58

DB 44 LIGLVVPYMEGLDSIEDLE---DQIEDKTIIVGIERGAGTVEAERALEDYG-KEDWEVI 99

QY 59 PSSSVMLASLSSIKRNEWILVPLWKPWF 90

DB 100 ASSGAMVTELRQAINNEEDIIVTGTPHWKF 131

RESULT 9

US-10-501-282-1796
 ; Sequence 1796, Application US/10501282
 ; Publication No. US20050203280A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MCMICHAEL, JOHN CALHOUN
 ; APPLICANT: ZAGURSKY, ROBERT JOHN
 ; APPLICANT: RUSSELL, DAVID PARRISH
 ; APPLICANT: FLETCHER, LEAH DIANE
 ; TITLE OF INVENTION: ALLOTOCOCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
 ; FILE REFERENCE: AM100780 L2
 ; CURRENT APPLICATION NUMBER: US/10/501,282
 ; CURRENT FILING DATE: 2004-07-09
 ; PRIOR APPLICATION NUMBER: 60/333,777
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: 60/426,742
 ; PRIOR FILING DATE: 2002-11-18

```
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1796
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
US-10-501-282-1796

Query Match          32.6%; Score 153; DB 5; Length 334;
Best Local Similarity 39.1%; Pred. No. 1.6e-09;
Matches 36; Conservative 19; Mismatches 31; Indels 6; Gaps 4;

QY 1 LQGFVVPVYP-SSISELKGDKFKNK-MIGIDAGAGTQIVTQALNYYGLSKEYELV 58
Db 105 LGLVVPFYNGLDSIEDLE---DQIEDKTIIGIERGAGTVEAERALEDYG-KEDWEVI 160

QY 59 PSSESVMLASLDSSIKRNEWILVPLWKPHWF 90
Db 161 ASSSGAMVTELRAQINNEEDIIVTGWTPHKF 192

RESULT 10
US-10-501-282-1798
; Sequence 1798, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1798
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
US-10-501-282-1798

Query Match          32.6%; Score 153; DB 5; Length 352;
Best Local Similarity 39.1%; Pred. No. 1.7e-09;
Matches 36; Conservative 19; Mismatches 31; Indels 6; Gaps 4;

QY 1 LQGFVVPVYP-SSISELKGDKFKNK-MIGIDAGAGTQIVTQALNYYGLSKEYELV 58
Db 123 LGLVVPFYNGLDSIEDLE---DQIEDKTIIGIERGAGTVEAERALEDYG-KEDWEVI 178

QY 59 PSSESVMLASLDSSIKRNEWILVPLWKPHWF 90
Db 179 ASSSGAMVTELRAQINNEEDIIVTGWTPHKF 210

RESULT 11
US-10-501-282-1792
; Sequence 1792, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
```

```
; TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1792
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Alloiococcus otitidis
US-10-501-282-1792

Query Match          32.4%; Score 152.5; DB 5; Length 320;
Best Local Similarity 36.4%; Pred. No. 1.7e-09;
Matches 32; Conservative 17; Mismatches 36; Indels 3; Gaps 1;

QY 3 GFVVPSYVPISSISELKGDKFKNKMGIDAGAGTQIVTQALNYYGLSKEYELVPSSE 62
Db 122 GLGVPTYMEVDSIADLT--DEAEQTITIGIEAGAGVVGQALEDYDNLSDWTVTATSSS 178

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWF 90
Db 179 GAMVTELETAIINNEEDIIVTAWSPHKF 206

RESULT 12
US-09-815-242-5213
; Sequence 5213, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlseen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5213
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5213

Query Match          29.4%; Score 138; DB 3; Length 312;
```



```
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49743
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-282-122A-49743

Query Match      27.0%; Score 127; DB 4; Length 380;
Best Local Similarity 30.3%; Pred. No. 2.5e-06;
Matches 27; Conservative 24; Mismatches 32; Indels 6; Gaps 3;

QY      4 FVVPYSY---VPISSISELKGKGFKNKMGIDAGAGTQIVTEQ--ALNYYGLSKEYELV 58
Db      184 FAVPSVEYDAGLKTADIAKHRDQLDGLGKIYGIPEGSSANAAIQKWTASNOFGL-RGFKLI 242
QY      59 PSSESVMLASLDSSIKRNEWILVPLWKPH 87
Db      243 ESSEAGMLVSDRAIREKKWVFLGWEPH 271
```

Search completed: January 24, 2006, 20:54:29
Job time : 16.9938 secs

RESULT 2
US-10-942-072-13
; Sequence 13, Application US/10942072
; Publication No. US20050250721A1
; GENERAL INFORMATION:
; APPLICANT: Hammon, H. K.
; APPLICANT: Insel, P. A.
; APPLICANT: Ping, P.
; APPLICANT: Post, S. R.
; APPLICANT: Gao, M.
; TITLE OF INVENTION: GENE THERAPY FOR CONGESTIVE HEART
; TITLE OF INVENTION: FAILURE
; FILE REFERENCE: 220002056723

	Db	742	YTLMPSPDVL	753	
			RESULT 4		
			US-10-518-341-1		
			/ Sequence 1, Application US/10518341		
			/ Publication No. US20050249744A1		
			/ GENERAL INFORMATION:		
			/ APPLICANT: VAN ELS, Cecile Antoinette Carola Maria		
			/ APPLICANT: BOOG, Marie-Claire Josefa Pia		
			/ TITLE OF INVENTION: MHC class II haplotype specific immunodominancy of		
			/ TITLE OF INVENTION: peptides derived from RSV fusion (F) or attachment		
			/ FILE REFERENCE: (G) proteins		
			/ FILE REFERENCE: 28902.0014		
			/ CURRENT APPLICATION NUMBER: US/10/518,341		
			/ CURRENT FILING DATE: 2004-12-17		
			/ PRIOR APPLICATION NUMBER: PCT/NL03/00454		
			/ PRIOR FILING DATE: 2003-06-20		
			/ PRIOR APPLICATION NUMBER: EP 02077461.8		
			/ PRIOR FILING DATE: 2002-06-20		
			/ NUMBER OF SEQ ID NOS: 29		
			/ SOFTWARE: PatentIn version 3.1		
			/ SEQ ID NO 1		
			/ LENGTH: 574		
			/ TYPE: PRP		
			/ ORGANISM: respiratory syncytial virus		
			US-10-518-341-1		
			Query Match	12.8%; Score 60; DB 6; Length 574;	
			Best Local Similarity	20.0%; Pred. No. 21;	
			Matches	16; Conservative 25; Mismatches 29; Indels 10; Gaps 3;	
			Qy	3 GFVPSYPPISSISBLKGGDKFKNMIG-----IDAGAGTQIVTEQAL---NYGLSKE 54	
			Db	145 GSATASGVAVSKVLHLEGEVNKIKSALLSTNKAVVSLNSGVSLTSKVLDLKNY--IDKQ 202	
			Qy	55 YELVPSSBSVMLASLDSIHK 74	
			Db	203 LLPVNVKQSCSISNIETVIE 222	
			RESULT 5		
			US-11-022-562-214		
			/ Sequence 214, Application US/11022562		
			/ Publication No. US20050249742A1		
			/ GENERAL INFORMATION:		
			/ APPLICANT: Ruprecht, Ruth M.		
			/ APPLICANT: Shisong, Jiang		
			/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING		
			/ TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE		
			/ FILE REFERENCE: DFN-043CN		
			/ CURRENT APPLICATION NUMBER: US/11/022,562		
			/ CURRENT FILING DATE: 2004-12-22		
			/ PRIOR APPLICATION NUMBER: PCT/US03/20322		
			/ PRIOR FILING DATE: 2003-06-27		
			/ PRIOR APPLICATION NUMBER: 60/392718		
			/ PRIOR FILING DATE: 2002-06-27		
			/ NUMBER OF SEQ ID NOS: 340		
			/ SOFTWARE: FastSeq for Windows Version 4.0		
			/ SEQ ID NO 214		
			/ LENGTH: 574		
			/ TYPE: PRP		
			/ ORGANISM: Respiratory Syncytial Virus		
			US-11-022-562-214		
			Query Match	12.8%; Score 60; DB 7; Length 574;	
			Best Local Similarity	20.0%; Pred. No. 21;	
			Matches	16; Conservative 25; Mismatches 29; Indels 10; Gaps 3;	
			Qy	3 GFVPSYPPISSISBLKGGDKFKNMIG-----IDAGAGTQIVTEQAL---NYGLSKE 54	
			Db	145 GSATASGVAVSKVLHLEGEVNKIKSALLSTNKAVVSLNSGVSLTSKVLDLKNY--IDKQ 202	

Db	742	YTIMPSPTDVL	753
Db	742	YTIMPSPTDVL	753
Query Match	12.8%	Score 60;	DB 6; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	
Matches	16;	Conservative 25;	Mismatches 29; Indels 10; Gaps 3;
QY	3	GFVPSYVPISSISBLKGGKDFKXKMGIG	-----IDAGAGTQIVTEQAL---NYGLSKE 54
Db	145	GSATASGAVAVSKVLHLEGEVKNKIKSALLSTNKAVVLSNGSVLTSKVLDLKNY--IDKQ 202	
QY	55	YELVPSSESVMLASLDSIK 74	
Db	203	LLPIVKNQSCSISNIETVIE 222	
Query Match	12.8%	Score 60;	DB 7; Length 574;
Best Local Similarity	20.0%	Pred. No. 21;	

Query Match 12.3%; Score 58; DB 7; Length 318;
Best Local Similarity 19.8%; Pred. No. 17;
Matches 17; Conservative 20; Mismatches 33; Indels


```
; APPLICANT: Insel, P. A.
; APPLICANT: Ping, P.
; APPLICANT: Post, S. R.
; APPLICANT: Gao, M.
; TITLE OF INVENTION: GENE THERAPY FOR CONGESTIVE HEART
; FILE REFERENCE: 22002056723
; CURRENT APPLICATION NUMBER: US/10/942,072
; PRIOR FILING DATE: 2004-09-14
; PRIOR APPLICATION NUMBER: US/09/750,240
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 09/472,667
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: US 09/008,097
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: US 08/924,757
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/048,933
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: US 08/708,661
; PRIOR FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-942-072-11
```

```
Query Match      12.2%; Score 57.5; DB 6; Length 1168;
Best Local Similarity 25.8%; Pred. No. 1e+02;
Matches 23; Conservative 9; Mismatches 26; Indels 31; Gaps 3;

QY 32 IDAG--AGTQIVTEQALNYGLSKEYELVLP-----SSE 62
DB 505 MEAGGAGRIHTRATLQY--LNGDYVEVPGCGGERNAYLKEQHETFIILGASQKRKEE 562

QY 63 SVMLASLDSSIKRNEWILVPLMKPHWAFS 91
DB 563 KAMLAQLQRTANSMEGLMPRVVPDRAFS 591
```

```
RESULT 15
US-10-982-545-6
; Sequence 6, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: CIPHERGEN Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; CURRENT FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 6
; LENGTH: 1663
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: complement component C3 precursor
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(23)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (23)..(667)
; OTHER INFORMATION: complement component C3 beta-chain
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (672)..(1663)
; OTHER INFORMATION: complement component C3 alpha-chain
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (672)..(748)
; OTHER INFORMATION: C3a anaphylatoxin peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (672)..(747)
; OTHER INFORMATION: C3a anaphylatoxin peptide
; OTHER INFORMATION: desArg peptide
US-10-982-545-6
```

```
Query Match      12.2%; Score 57.5; DB 6; Length 1663;
Best Local Similarity 27.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 9; Mismatches 15; Indels 23; Gaps 2;

QY 21 KGDRFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSESVMLASLDSS---IKRNE 77
DB 97 KSEGRKNKFVTVQATFGTQVV-----EKVVLVSLQSGYLFIQTDK 136

QY 78 WILVP 82
DB 137 TIYTP 141
```

Search completed: January 24, 2006, 20:55:34
Job time : 2.45643 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 3.45228 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-6
Perfect score: 470
Sequence: 1 LQGFVPSYVPISSELK.....SIKNEWILVPLKPHWAFS 91

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	468	99.6	290	2 H70117	glycine betaine, L
2	193	41.1	285	2 H83241	probable glycine b
3	186.5	39.7	573	2 C86806	hypothetical prote
4	176	37.4	292	2 AI3042	hypothetical prote
5	176	37.4	304	2 C98243	hypothetical prote
6	175.5	37.3	287	2 AE3578	glycine betaine/l-
7	138	29.4	312	2 H82972	hypothetical prote
8	127	27.0	312	2 C82374	hypothetical prote
9	127	27.0	312	2 D95891	probable glycine-b
10	126	26.8	307	2 D83642	hypothetical prote
11	124.5	26.5	233	2 I40537	glycine betaine AB
12	124.5	26.5	300	2 AF1559	glycine betaine AB
13	124.5	26.5	300	2 AH1201	glycine betaine AB
14	108	23.0	308	2 AE2829	hypothetical prote
15	108	23.0	337	2 C97607	hypothetical prote
16	104	22.1	326	2 AC3307	glycine betaine/l-
17	101	21.5	317	2 AC3128	conserved hypotet
18	101	21.5	363	2 F98159	hypothetical prote
19	88	18.7	314	2 H83321	conserved hypotet
20	82.5	17.6	265	2 E97633	hypothetical prote
21	82.5	17.6	317	2 AH2856	hypothetical prote
22	80	17.0	331	2 AC0842	glycine betaine-bi
23	76.5	16.3	903	2 F82080	preprotein translo
24	76	16.2	322	2 E83009	probable binding p
25	75	16.2	327	2 H59094	hypothetical prote
26	74	15.7	319	2 F84966	thioredoxin-disulf
27	73.5	15.6	687	2 T04327	probable serine/th
28	73	15.5	334	2 AG0322	glycine betaine-bi
29	72	15.3	445	2 G70371	UDP-MURNAC-pentape

30	71.5	15.2	356	2 JX0141	cytochrome-c oxida
31	69.5	14.8	214	2 S46476	cysteine proteinas
32	69	14.7	467	2 S61141	probable membrane
33	68.5	14.6	345	2 AG3186	hypothetical prote
34	68	14.5	299	2 AI1208	metal binding prot
35	67.5	14.4	1182	2 E97286	DNA-dependent RNA
36	67.5	14.4	1199	2 G69698	RNA polymerase (be
37	67.5	14.4	1393	2 B86501	RNA polymerase bet
38	67.5	14.4	1393	2 E72122	DNA polymerase bet
39	67.5	14.4	1396	2 F81686	DNA-directed RNA p
40	67.5	14.4	1396	2 G71529	DNA-directed RNA p
41	67.5	14.4	1397	2 B81548	DNA-directed RNA p
42	67	14.3	562	2 AD0304	probable hemolysin
43	67	14.3	1005	2 T31333	beta-galactosidase
44	66	14.0	270	2 A82245	periplasmic bindin
45	66	14.0	323	2 B83215	conserved hypotet

ALIGNMENTS

RESULT 1

H70117
glycine betaine, L-proline ABC transporter, glycine/betaine/L-proline-binding protein (L
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: H70117
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whit-
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: H70117
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-290 <KLE>
A:Cross-references: UNIPROT:O51169; UNIPARC:UPI000005737B; GB:AE001125; GB:AE000783; NIT
A:Experimental source: strain B31

Query Match	99.6%	Score	468	DB 2	Length	290			
Best Local Similarity	98.9%	Pred. No.	1.1e-41						
Matches	90	Conservative	1	Mismatches	0	Indels	0	Gaps	0
QY	1	LQGFVPSYVPISSELKGGKDKFNKMIGIDAGAGTQIVTEQALNYGLSKVELYPS	60						
	:	:	:	:	:	:	:	:	:
Db	109	IQGFVPSYVPISSELKGGKDKFNKMIGIDAGAGTQIVTEQALNYGLSKVELYPS	168						
	:	:	:	:	:	:	:	:	:
QY	61	SESVMASLSDSIKRNWILVPLKPHWAFS	91						
	:	:	:	:	:	:	:	:	:
Db	169	SESVMASLSDSIKRNWILVPLKPHWAFS	199						
	:	:	:	:	:	:	:	:	:

RESULT 2

H83241
probable glycine betaine-binding protein precursor PA3236 [imported] - Pseudomonas aeru
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83241
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lm,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-285 <STO>
A:Cross-references: UNIPROT:Q9HZ04; UNIPARC:UPI00000C5968; GB:AE004746; GB:AE004091; NIT
A:Experimental source: strain PAO1
C:Genetics:

A:Gene: PA3236

Query Match 41.1%; Score 193; DB 2; Length 285;
Best Local Similarity 42.7%; Pred. No. 9.7e-13;
Matches 38; Conservative 20; Mismatches 29; Indels 2; Gaps 1;

QY 3 GFVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYELVPSS 62
DB 116 GLIVPEYUSANSIADLQAKDAFGGRVVGIDAGAGWIKTDEAIKQYGL--DYKLVAASG 173

QY 63 SVMLASLSSIKRNEWILVPLWKPWFPS 91
DB 174 SGMIAELTAENEKRPVVVTGWIPIHMF 202

RESULT 3

C86806
hypothetical protein busAB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86806
R:Solotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86806
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <STO>
A:Cross-references: UNIPROT:O9KIP6; UNIPARC:UPI00000C611D; GB:AE005176; PID:g12724442; F
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: busAB

Query Match 39.7%; Score 186.5; DB 2; Length 573;
Best Local Similarity 40.4%; Pred. No. 1.1e-11;
Matches 36; Conservative 21; Mismatches 29; Indels 3; Gaps 1;

QY 3 GFVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYELVPSS 62
DB 406 GFVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYELVPSS 462

QY 63 SVMLASLSSIKRNEWILVPLWKPWFPS 91
DB 463 GATVALGEAIKQKDIVTGSPPHMFN 491

RESULT 4

A13042
hypothetical protein Atu3957 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: A13042
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: A13042
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <KUR>
A:Cross-references: UNIPROT:Q8U8Y1; UNIPARC:UPI00001648AD; GB:AE008689; PIDN:AAL44759.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3957
A:Map position: linear chromosome

Query Match 37.4%; Score 176; DB 2; Length 292;
Best Local Similarity 41.4%; Pred. No. 6.2e-11;

Matches 36; Conservative 14; Mismatches 35; Indels 2; Gaps 1;

QY 2 QGVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYELVPSS 61
DB 119 QAIAPVKVTTIDSMQDLNDNADKFGGKIIGIEPGSLMRDAGAKADYGL--KLQLVGS 176

QY 62 ESMVLASLSSIKRNEWILVPLWKPWFPS 88
DB 177 TAAMTAALKSAIDRKEWAVTVWPEPSW 203

RESULT 5

C98243
hypothetical protein AGR_L1793 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: C98243
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: C98243
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <KUR>
A:Cross-references: UNIPROT:Q8U8Y1; UNIPARC:UPI00000D23BB; GB:AE007870; PIDN:AAK89469.1;
C:Genetics:
A:Gene: AGR_L1793
A:Map position: linear chromosome

Query Match 37.4%; Score 176; DB 2; Length 304;
Best Local Similarity 41.4%; Pred. No. 6.5e-11;
Matches 36; Conservative 14; Mismatches 35; Indels 2; Gaps 1;

QY 2 QGVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYELVPSS 61
DB 131 QAIAPVKVTTIDSMQDLNDNADKFGGKIIGIEPGSLMRDAGAKADYGL--KLQLVGS 188

QY 62 ESMVLASLSSIKRNEWILVPLWKPWFPS 88
DB 189 TAAMTAALKSAIDRKEWAVTVWPEPSW 215

RESULT 6

AE3578
glycine betaine/l-proline-binding protein proX [imported] - Brucella melitensis (strain
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AE3578
R:DelVecchio, V.G.; Kapural, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goitsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A:Reference number: AB3252; PMID:11756688
A:Accession: AE3578
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-287 <KUR>
A:Cross-references: UNIPROT:Q8YCH8; UNIPARC:UPI0000058515; GB:AE008918; PIDN:AAL53792.1
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME110550
A:Map position: II

Query Match 37.3%; Score 175.5; DB 2; Length 287;
Best Local Similarity 38.7%; Pred. No. 6.9e-11;
Matches 36; Conservative 21; Mismatches 29; Indels 7; Gaps 3;

QY 3 GFVVPSPVPISSISELKGDKGKFNKMGIDAGAGTQIVTEQALNYGLSKEYEL 57
DB 114 GMIVPDYTPESVSSIEDLKKPEVREKLGETQIGDPGAGLTRLQAEAIKKYGL--DYKL 171

```
QY 58 VPSSSVMLASLDSSIKRNEWILVPLWKPHWAF 90
Db 172 NISSEAMLTVDRAIRSEGMFVATSWSPHMF 204

RESULT 7
H82972
hypothetical protein PA5388 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H82972
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H82972
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <STO>
A:Cross-references: UNIPROT:Q9HTH6; UNIPARC:UPI00000C5FD9; GB:AE004951; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5388

Query Match 29.4%; Score 138; DB 2; Length 312;
Best Local Similarity 33.7%; Pred. No. 6.8e-07;
Matches 30; Conservative 24; Mismatches 29; Indels 6; Gaps 3;

QY 4 FVPSYVYP---ISSISELKGKDKFKNMIGIDAGAGTQIVTEQAL--NYGLSKEYELV 58
Db 117 YAVPSYVAGGLKTPADARFKDKLGRIYALEPGSGSNRITRKIMIDNRFGEL-KGFQLV 175

QY 59 PSSESVMLASLDSSIKRNEWILVPLWKPH 87
Db 176 ESSEAGMLTAVKRAIKRQWVFFGCKPH 204

RESULT 8
C82974
hypothetical protein PA5378 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: C82974
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C82974
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <STO>
A:Cross-references: UNIPROT:Q9HTI6; UNIPARC:UPI00000C5FCF; GB:AE004950; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5378

Query Match 27.0%; Score 127; DB 2; Length 312;
Best Local Similarity 34.5%; Pred. No. 9.8e-06;
Matches 30; Conservative 20; Mismatches 31; Indels 6; Gaps 3;

QY 6 VPSYV---PISSISELKGKDKFKNMIGIDAGAGTQIVTEQAL--NYGLSKEYELVPS 60
Db 117 VPQYVDGLRSFADIAKPSDKLGNKIYGLIEPGNGNRVAOSMIDKNAPELGK-PKLVES 175

QY 61 SESVVMLASLDSSIKRNEWILVPLWKPH 87
Db 176 SEAGMLSQVQAIRNRQWVFFLGWEPH 202
```

```
RESULT 9
D95891
probable glycine-betaine-binding protein [imported] - Sinorhizobium meliloti (strain 102
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: D95891
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95891
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <KUR>
A:Cross-references: UNIPROT:Q92WE4; UNIPARC:UPI00000CB4FA; GB:AL591985; PIDN:CAC48796.1;
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaucht, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wellis, D.H.; Wong, K.; Yeh, K.
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SM520410
A:Genome: plasmid

Query Match 27.0%; Score 127; DB 2; Length 312;
Best Local Similarity 37.1%; Pred. No. 9.8e-06;
Matches 33; Conservative 16; Mismatches 33; Indels 6; Gaps 3;

QY 4 FVPSYVYP---ISSISELKGKDKFKNMIGIDAGAGTQIVTEQALNY--YGLSKEYELV 58
Db 115 FAVPKYVADAGVKDISDLQKFPDKFRKIYGLIEPGNGNRMLDMINKGDFGLT-GWELV 173

QY 59 PSSESVMLASLDSSIKRNEWILVPLWKPH 87
Db 174 ESSEQGLAVERATKDDQWIVFLGWAPH 202

RESULT 10
D83642
hypothetical protein PA0030 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: D83642
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83642
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <STO>
A:Cross-references: UNIPROT:Q9I7A0; UNIPARC:UPI00000C4EF4; GB:AE004442; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0030

Query Match 26.8%; Score 126; DB 2; Length 307;
Best Local Similarity 34.5%; Pred. No. 1.2e-05;
Matches 30; Conservative 16; Mismatches 35; Indels 6; Gaps 3;

QY 6 VPSYV---PISSISELKGKDKFKNMIGIDAGAGTQIVTEQAL--NYGLSKEYELVPS 60
Db 117 VPRIYWDAGVHRFEDLAAQGOGRFNKLYIGISGAPANQSIQKIMIDANQFGLG-DWKLVES 175

QY 61 SESVVMLASLDSSIKRNEWILVPLWKPH 87
```

Db 176 SEQAMLAELGRAEKQRMLVFLGWTPH 202

RESULT 11

glycine betaine ABC transporter (glycine betaine-binding protein) opuAC precursor - Bacillus subtilis

C:Species: *Bacillus subtilis*

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C:Accession: I40537, F69669

R:Kempf, B.; Bremer, E.

J. Biol. Chem. 270, 16701-16713, 1995

A:Title: OpuA, an osmotically regulated protein-dependent transport system for betaine

A:Reference number: A57322; MUID:95348093; PMID:7622480

A:Accession: I40537

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-293 <RES>

A:Cross-references: UNIPROT:P46922; UNIPARC:UPI000005FED3; EMBL:U17292; NID:9984802; PID:140537

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chedoke, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maugda, S.; Maueel, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69669

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-293 <KUN>

A:Cross-references: UNIPARC:UPI000005FED3; GB:Z99105; GB:AL009126; NID:92632457; PIDN:CA

A:Experimental source: strain 168

C:Genetics:

A:Gene: opuAC

Query Match 26.5%; Score 124.5; DB 2; Length 293;

Best Local Similarity 32.9%; Pred. No. 1.7e-05;

Matches 26; Conservative 18; Mismatches 34; Indels 1; Gaps 1;

QY 13 SSTSELKRGDKPKNMIGIDAGAGTQIVTEQALNYIGLSKEVELVPSSVSMASLDSS 72

Db 23 SENDENASAEQVNTIIGIDPGSGIMSLTRAMKDYDLN-DWTLISASSAAMTATLKS 81

QY 73 IKRNEWILPLMKPHWAFS 91

Db 82 YDRKKPIITGTWTPHMF 100

RESULT 12

AF1559

glycine betaine ABC transporters (glycine betaine-binding protein) homolog gbuC [imported]

C:Species: *Listeria innocua*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AF1559

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AF1559

A:Status: preliminary

A:Molecule type: DNA

Query Match 26.5%; Score 124.5; DB 2; Length 300;

Best Local Similarity 35.1%; Pred. No. 1.7e-05;

Matches 26; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 19 KKGDKFKNMIGIDAGAGTQIVTEQALNYIGLSKE-YELVPSSVSMASLDSSIKRNE 77

Db 32 KDLGSEQINYTTIGIDAGAGIMLATQNAIKDYHLDDNNWQLQTSSTAAMTSTLQKAMKDKR 91

QY 78 WILVPLMKPHWAFS 91

Db 92 PIVVTGWTTPHWNFT 105

RESULT 13

AH1201

glycine betaine ABC transporters (glycine betaine-binding protein) homolog gbuC [imported]

C:Species: *Listeria monocytogenes*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AH1201

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1201

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-300 <GLA>

A:Cross-references: UNIPROT:Q8Y898; UNIPARC:UPI00000CF12A; GB:NC_003210; PIDN:CAC99094.1

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: gbuC

Query Match 26.5%; Score 124.5; DB 2; Length 300;

Best Local Similarity 35.1%; Pred. No. 1.7e-05;

Matches 26; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 19 KKGDKFKNMIGIDAGAGTQIVTEQALNYIGLSKE-YELVPSSVSMASLDSSIKRNE 77

Db 32 KDLGSEQINYTTIGIDAGAGIMLATQNAIKDYHLDDNNWQLQTSSTAAMTSTLQKAMKDKR 91

QY 78 WILVPLMKPHWAFS 91

Db 92 PIVVTGWTTPHWNFT 105

RESULT 14

AE2829

hypothetical protein Atu2060 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C:Species: *Agrobacterium tumefaciens*

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

C:Accession: AE2829

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, J.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AE2829

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <KUR>

A:Residues: 1-300 <GLA>

A:Cross-references: UNIPROT:Q92D06; UNIPARC:UPI00000CC423; GB:AL592022; PIDN:CAC96246.1;

A:Experimental source: strain Clip11262

C:Genetics:

A:Gene: gbuC

Query Match 26.5%; Score 124.5; DB 2; Length 300;

Best Local Similarity 35.1%; Pred. No. 1.7e-05;

Matches 26; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 19 KKGDKFKNMIGIDAGAGTQIVTEQALNYIGLSKE-YELVPSSVSMASLDSSIKRNE 77

Db 32 KDLGSEQINYTTIGIDAGAGIMLATQNAIKDYHLDDNNWQLQTSSTAAMTSTLQKAMKDKR 91

QY 78 WILVPLMKPHWAFS 91

Db 92 PIVVTGWTTPHWNFT 105

RESULT 13

AH1201

glycine betaine ABC transporters (glycine betaine-binding protein) homolog gbuC [imported]

C:Species: *Listeria monocytogenes*

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004

C:Accession: AH1201

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A:Title: Comparative genomics of *Listeria* species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1201

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-300 <GLA>

A:Cross-references: UNIPROT:Q8Y898; UNIPARC:UPI00000CF12A; GB:NC_003210; PIDN:CAC99094.1

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: gbuC

Query Match 26.5%; Score 124.5; DB 2; Length 300;

Best Local Similarity 35.1%; Pred. No. 1.7e-05;

Matches 26; Conservative 16; Mismatches 31; Indels 1; Gaps 1;

QY 19 KKGDKFKNMIGIDAGAGTQIVTEQALNYIGLSKE-YELVPSSVSMASLDSSIKRNE 77

Db 32 KDLGSEQINYTTIGIDAGAGIMLATQNAIKDYHLDDNNWQLQTSSTAAMTSTLQKAMKDKR 91

QY 78 WILVPLMKPHWAFS 91

Db 92 PIVVTGWTTPHWNFT 105

RESULT 14

AE2829

hypothetical protein Atu2060 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)

C:Species: *Agrobacterium tumefaciens*

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

C:Accession: AE2829

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, J.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AE2829

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <KUR>

A:Cross-references: UNIPROT:Q8UDR1; UNIPARC:UPI000016469D; GB:AE008688; PIDN:AAL43051.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2060
A:Map position: circular chromosome

Query Match 23.0%; Score 108; DB 2; Length 308;
Best Local Similarity 34.4%; Pred. No. 0.00097;
Matches 31; Conservative 16; Mismatches 33; Indels 10; Gaps 4;
QY 4 FVPPSVV---PISSISELKGDKFKNMIGIDAGAG---TQIVTEQALNYYGLSKEYEL 57
DB 113 FAVPTVWEAGVKDFSDLQKPADKDRKLYGIEPGSNQLMLDAVKDPAL---GL-KDWEV 168
QY 58 VPSSSVMLASLDSSIKRNEWILVPLWKPH 87
DB 169 VESSEQGLSQVAREFNKTKFIVFGWAPH 198

RESULT 15
C97607
hypoetical protein AGR_C 3730 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: C97607
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: C97607
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
C:Cross-references: UNIPROT:Q8UDR1; UNIPARC:UPI00000D1D7F; GB:AE007869; PIDN:AAK87812.1;
A:Gene: AGR_C 3730
A:Map position: circular chromosome

Query Match 23.0%; Score 108; DB 2; Length 337;
Best Local Similarity 34.4%; Pred. No. 0.0011;
Matches 31; Conservative 16; Mismatches 33; Indels 10; Gaps 4;
QY 4 FVPPSVV---PISSISELKGDKFKNMIGIDAGAG---TQIVTEQALNYYGLSKEYEL 57
DB 142 FAVPTVWEAGVKDFSDLQKPADKDRKLYGIEPGSNQLMLDAVKDPAL---GL-KDWEV 197
QY 58 VPSSSVMLASLDSSIKRNEWILVPLWKPH 87
DB 198 VESSEQGLSQVAREFNKTKFIVFGWAPH 227

Search completed: January 24, 2006, 19:59:01
Job time : 4.45228 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 19.5539 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-6

Perfect score: 470

Sequence: 1 LQGFVVPYSYPISSISLKG.....SIKNEWILVPLMKPHWAFS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	99.6	290	051169 BORBU	051169 borrelia bu
2	459	97.7	294	066219 BORBU	066219 borrelia bu
3	218	46.4	277	Q8A6X6 BACTN	Q8A6X6 bacteroides
4	218	46.4	315	Q8TNY0 METAC	Q8TNY0 methanosarc
5	216.5	46.1	285	Q729Q3 DESVH	Q729Q3 desulfovibr
6	211.5	45.0	284	Q64R03 BACFR	Q64R03 bacteroides
7	210.5	44.8	284	Q5LAK8 BACFN	Q5LAK8 bacteroides
8	199	42.3	871	Q9RJ81 STRCO	Q9RJ81 streptomyce
9	194	41.3	287	Q63M02 BURPS	Q63M02 burkholderi
10	194	41.3	287	Q62DH6 BURMA	Q62DH6 burkholderi
11	193	41.1	285	Q9HZ04 PSEAE	Q9HZ04 pseudomonas
12	193	41.1	287	Q4LV54 9BURK	Q4LV54 burkholderi
13	192	40.9	871	Q828F3 STRAM	Q828F3 streptomyce
14	187.5	39.9	283	Q4KFZ7 PSEF5	Q4KFZ7 pseudomonas
15	186.5	39.7	573	Q9KIF6 9LACT	Q9KIF6 lactococcus
16	186.5	39.7	573	Q7DAU8 LACTA	Q7DAU8 lactococcus
17	183.5	39.0	306	Q8U4S3 METMA	Q8U4S3 methanosarc
18	180.5	38.4	573	Q9RQ05 9LACT	Q9RQ05 lactococcus
19	178	37.9	287	Q8M34 PSEPK	Q8M34 pseudomonas
20	177.5	37.8	285	Q5WCG0 BACSK	Q5WCG0 bacillus cl
21	176.5	37.6	288	Q8FVT7 BRUSU	Q8FVT7 brucella su
22	176	37.4	292	Q8U8Y1 AGRT5	Q8U8Y1 agrobacteri
23	176	37.4	304	Q7CT01 AGRT5	Q7CT01 agrobacteri
24	175.5	37.3	287	Q8YCH8 BRUME	Q8YCH8 brucella me
25	175.5	37.3	288	Q578M1 BRUAB	Q578M1 brucella ab
26	174	37.0	302	Q7W819 PROMM	Q7W819 prochloroco
27	172.5	36.7	285	Q63AG1 BACCZ	Q63AG1 bacillus ce
28	169	36.0	321	Q7U4Z5 SYNPK	Q7U4Z5 synecococc
29	169	36.0	328	Q886F7 PSESM	Q886F7 pseudomonas
30	168.5	35.9	285	Q6HHV0 BACHK	Q6HHV0 bacillus th
31	168.5	35.9	285	Q81PL5 BACAN	Q81PL5 bacillus an

32	168	35.7	289	2	Q8XTP2_RALSO	Q8XTP2 ralstonia s
33	166	35.3	284	2	Q4ZPY3_PSESY	Q4ZPY3 pseudomonas
34	166	35.3	295	2	Q6LYW9_METWP	Q6LYW9 methanococc
35	166	35.3	298	2	Q891M9_CLOTE	Q891M9 clostridium
36	165.5	35.2	285	2	Q4MKI4_BACCE	Q4MKI4 bacillus ce
37	164	34.9	282	2	Q5WJD4_BACSK	Q5WJD4 bacillus cl
38	163.5	34.8	285	2	Q81CH2_BACCR	Q81CH2 bacillus ce
39	158.5	33.7	569	2	Q830X6_ENTFA	Q830X6 enterococcu
40	157	33.4	575	2	Q9A1M9_STRPY	Q9A1M9 streptococc
41	156	33.2	561	2	Q8K8R4_STRP3	Q8K8R4 streptococc
42	156	33.2	575	2	Q879N8_STRP3	Q879N8 streptococc
43	156	33.2	575	2	Q5XE24_STRP6	Q5XE24 streptococc
44	156	33.2	575	2	Q8P2S5_STRP8	Q8P2S5 streptococc
45	153	32.6	308	2	Q98HR5_RHILO	Q98HR5 rhizobium 1

ALIGNMENTS

RESULT 1

ID 051169 BORBU PRELIMINARY; PRT; 290 AA.
AC 051169
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Glycine betaine, L-proline ABC transporter, glycine/betaine/L-proline-binding protein (ProX)
DE binding protein (ProX)
GN OrderedLocusNames=BB0144;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_taxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K., Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D., Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J., Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.F., Utterback T.R., Matthey L., McDonald L.A., Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K., Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RA "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL Nature 390:580-586(1997).
RL EMBL; AB001125; AAC6525.1; -; Genomic_DNA.
DR PIR; H70117; H70117.
DR TIGR; BB0144; -;
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 290 AA; 33263 MW; 43779CF0911FBBD5 CRC64;

Query Match 99.6%; Score 468; DB 2; Length 290;
Best Local Similarity 98.9%; Pred. No. 2.3e-40;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	LQGFVVPYSYPISSISLKGDKFKPKMKMGIDAGAGTQIVTEQALNYGLSKVELVPS	60
Db	109	LQGFVVPYSYPISSISLKGDKFKPKMKMGIDAGAGTQIVTEQALNYGLSKVELVPS	168
QY	61	SSSVMLASLDSIKRNEWILVPLMKPHWAFS	91
Db	169	SSSVMLASLDSIKRNEWILVPLMKPHWAFS	199

RESULT 2

```
Q662L9 BORGA
ID Q662L9_BORGA PRELIMINARY; PRT; 294 AA.
AC Q662L9_
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Glycine betaine, L-proline ABC transporter, binding protein.
GN Name=proX; OrderedLocNames=BG0144;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Pb1;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechdel U., Wilske B., Stuehn J., Platzer M.;
RA "Comparative analysis of the Borrelia garinii genome.";
RT Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; CP000013; RAU07002.1; -; Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 294 AA; 33680 MW; AFA81026194ABE12 CRC64;

Query Match 97.7%; Score 459; DB 2; Length 294;
Best Local Similarity 95.6%; Pred. No. 2e-39;
Matches 87; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LOGFVPSVYPISSISLKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPS 60
DB 113 LOGFVPSVYPISSISLKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPS 172
QY 61 SSVMLASLDSSIKRNEWILVPLWKPHWAFS 91
DB 173 SSVMLASLDSSAIKRNWILVPLWKPHWAFS 203

RESULT 3
ID Q6AX6_BACTN PRELIMINARY; PRT; 277 AA.
AC Q6AX6_
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glycine betaine-binding protein.
GN OrderedLocNames=BT1749;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Hmrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RA "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; A8016933; AA076856.1; -; Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 277 AA; 31056 MW; B020C694CE182A21A CRC64;

Query Match 46.4%; Score 218; DB 2; Length 277;
Best Local Similarity 47.7%; Pred. No. 2.7e-14;
Matches 45; Conservative 9; Mismatches 33; Indels 2; Gaps 1;

QY 3 GFVVPSPYPISSISLKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPSSE 62
DB 142 GLVVPYVITDSIEELNKEKDFNGQIIGIDFGAGIMQASEITADYDL--DMELVSGSS 199
QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAFS 91
DB 200 AMVTSLSKSDSEEWVTVTLNSPWFN 228

RESULT 5
ID Q729Q3_DESVH PRELIMINARY; PRT; 285 AA.
AC Q729Q3_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Glycine/betaine/L-proline ABC transporter, periplasmic-binding

Matches 42; Conservative 16; Mismatches 28; Indels 2; Gaps 1;

QY 3 GFVVPSPYPISSISLKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPSSE 62
DB 105 GLVVPDYVYITINSIELNAEKRFSGQIVGIDAGAGIMKATQAIKDYGL--DYKLMTSSG 162
QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAF 90
DB 163 PAMTASLKKAIDKDDWIVVTGWTPHWPF 190

RESULT 4
Q8TNY0_METAC PRELIMINARY; PRT; 315 AA.
AC Q8TNY0_
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glycine betaine/L-proline ABC transporter, solute-binding
DE protein.
GN Name=proX; OrderedLocNames=MA2147;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Navlor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talanas J., Tirrell A., Ye W., A.M.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Guss A.M.,
RA Hedderich R., Ingran-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010901; AAM05545.1; -; Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 315 AA; 34613 MW; 262494BFC79649BA CRC64;

Query Match 46.4%; Score 218; DB 2; Length 315;
Best Local Similarity 50.6%; Pred. No. 3.1e-14;
Matches 45; Conservative 9; Mismatches 33; Indels 2; Gaps 1;
```

Query Match	45.0%;	Score	211.5;	DB 2;	Length	284;			
Best Local Similarity	46.8%;	Pred. NO.	1.7e-13;						
Matches	41;	Conservative	16;	Mismatches	30;	Indels	1;	Gaps	1;
QY	3	GFVVPVVPISSELKKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYEYELVPSSE	62						
DB	112	GLVVPQVVTIDSLGELAAHKQFSSEIVGIDAGAGIMKTTDKAIGDYGLD-GYKLLTSSS	170						
QY	63	SVMLASLDSSIKRNEWILVPLWKPHWAF	90						
DB	171	STWLASLQKAMEKEAWIVTGMTPHWAF	198						
<p>RESULT 7</p> <p>Q5LAK8_BACFN</p> <p>ID Q5LAK8_BACFN PRELIMINARY; PRT; 284 AA.</p> <p>AC Q5LAK8;</p> <p>DT 01-FEB-2005 (TrEMBLrel. 29, Created)</p> <p>DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)</p> <p>DE 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)</p> <p>DE Putative exported glycine betaine-binding protein.</p> <p>GN OrderedLocusNames=BF3170;</p> <p>OS Bacteroides fragilis (strain ATCC 25285 / NCTC 9343).</p> <p>OC Bacteria; Bacteroidetes; Bacteroidetes (class); Bacteroidales;</p> <p>OC Bacteroidaceae; Bacteroides.</p> <p>OC NCBI_TaxID=272559;</p> <p>RP [1]</p> <p>RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].</p> <p>RX PubMed=15746427; DOI=10.1126/science.1107008;</p> <p>RA Cardeno-Tarraga A.-M., Patrick S., Crossman L.C., Blakely G., Abratt V., Lennard N., Poxton I., Duerden B., Harris B., Quail M.A., Barron A., Clark L., Corton C., Doggett J., Holden M.T.G., Larke N., Lane A., Lord A., Norbertczak H., Ormond D., Price C., Rabinowitz E., Woodward J., Barrell B.G., Parkhill J.;</p> <p>RT "Extensive DNA inversions in the <i>B. fragilis</i> genome control variable gene expression.";</p> <p>RL Science 307:1463-1465(2005).</p> <p>LR EMBL; CR626927; CAH08865.1; -; Genomic DNA.</p> <p>DR InterPro; IPR007210; OpuAC_ABC_sub_bd.</p> <p>DR Pfam; PF04069; OpuAC; 1.</p> <p>KW Complete proteome.</p> <p>QY SEQUENCE 284 AA; 31962 MW; 4F34F719D44F0448 CRC64;</p>									
Query Match	44.8%;	Score	210.5;	DB 2;	Length	284;			
Best Local Similarity	46.6%;	Pred. NO.	1.7e-13;						
Matches	41;	Conservative	16;	Mismatches	30;	Indels	1;	Gaps	1;
QY	3	GFVVPVVPISSELKKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKYEYELVPSSE	62						
DB	112	GLVVPQVVTIDSLGELAAHKQFSSEIVGIDAGAGIMKTTDKAIGDYGLD-GYKLLTSSS	170						
QY	63	SVMLASLDSSIKRNEWILVPLWKPHWAF	90						
DB	171	STWLASLQKAMEKEAWIVTGMTPHWAF	198						
<p>RESULT 8</p> <p>Q9RJ81_STRCO</p> <p>ID Q9RJ81_STRCO PRELIMINARY; PRT; 871 AA.</p> <p>AC Q9RJ81;</p> <p>DT 01-MAY-2000 (TrEMBLrel. 13, Created)</p> <p>DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)</p> <p>DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)</p> <p>DE Glycine betaine transport system permease protein.</p> <p>GN OrderedLocusNames=SCO1620; ORFNames=SCI41.03c;</p> <p>OS Streptomyces coelicolor.</p> <p>OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;</p> <p>OC Streptomycineae; Streptomycetaceae; Streptomyces.</p> <p>OC NCBI_TaxID=1902;</p> <p>RP [1]</p> <p>RP NUCLEOTIDE SEQUENCE.</p> <p>RC STRAIN=A3(2) / M145;</p>									

```

RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidaigo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147 (2002).
CC -!- FUNCTION: Part of a binding-protein-dependent transport system.
CC Probably responsible for the translocation of the substrate across
CC the membrane (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the binding-protein-dependent transport
CC system permease family.
DR EMBL; AL939109; CAB59473.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF00528; BPD transp_1; 1.
DR Pfam; PF04069; OpuAC; 2.
DR PROSITE; PS0928; ABC TM1; 1.
DR Complete proteome; Transmembrane; Transport.
KW SEQUENCE 871 AA; 95028 MW; 4CC12B484AA43506 CRC64;

Query Match 42.3%; Score 199; DB 2; Length 871;
Best Local Similarity 46.6%; Pred. No. 9.3e-12;
Matches 41; Conservative 14; Mismatches 31; Indels 2; Gaps 2;

QY 6 VPSYVPISSISSELKGGKFKNMIGIDAGAG-TQIVTEQALNYGLSKYELVPSSES 63
DB 699 VFSYVKDVKSLADKGLDFGKIGIEFGEGMQLKNVLPVGLDEYEVVDGSGTP 758

QY 64 VMLASLDSSIKRNEWILVPLWKPHWAPS 91
DB 759 AMLAEKRALAKKEPVAVTLMSPHWAYS 786

RESULT 9
Q63MU2 BURPS
ID Q63MU2 BURPS PRELIMINARY; PRT; 287 AA.
AC Q63MU2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Putative exported protein.
GN OrderedLocNames=BP05056;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxId=28450;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Tibball R.W., Peacock S.J., Cerdeno-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crossett B., Davis P., Deshaizer D.,
RA Felwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tunapa S., Vesaratchavest M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;

```

```

RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245 (2004).
DR EMBL; BX571966; CAH38013.1; -; Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000583; GATase 2.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
DR PROSITE; PS00443; GATASE_TYPE_II; UNKNOWN_1.
DR Complete proteome.
KW SEQUENCE 287 AA; 30851 MW; EF3DFA0B95DC82AB CRC64;

Query Match 41.3%; Score 194; DB 2; Length 287;
Best Local Similarity 42.7%; Pred. No. 9e-12;
Matches 38; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 3 GFVPSYVPISSISSELKGGKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPSSE 62
DB 116 GLIVPAYVKAKTIADLNQAQKNDFAGRIVGIDAGAGWMKRTDDAIKAYGLN--YSLMPSSG 173

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAPS 91
DB 174 SAMTAELARSIHSNKAVVVTGWAPHWMPA 202

RESULT 10
Q62DH6 BURMA
ID Q62DH6 BURMA PRELIMINARY; PRT; 287 AA.
AC Q62DH6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE ABC transporter, periplasmic glycine/betaine-binding protein.
GN OrderedLocNames=BMAA0480;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxId=13373;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nieman W.C., Deshaizer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M.,
RA Daugherty S.C., Davidson T.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251 (2004).
DR EMBL; CP000011; AAU46649.1; -; Genomic_DNA.
DR TIGR; BMAA0480; -;
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000583; GATase 2.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
DR PROSITE; PS00443; GATASE_TYPE_II; UNKNOWN_1.
DR Complete proteome.
KW SEQUENCE 287 AA; 30851 MW; EF3DFA0B95DC82AB CRC64;

Query Match 41.3%; Score 194; DB 2; Length 287;
Best Local Similarity 42.7%; Pred. No. 9e-12;
Matches 38; Conservative 18; Mismatches 31; Indels 2; Gaps 1;

QY 3 GFVPSYVPISSISSELKGGKFKNMIGIDAGAGTQIVTEQALNYGLSKYELVPSSE 62
DB 116 GLIVPAYVKAKTIADLNQAQKNDFAGRIVGIDAGAGWMKRTDDAIKAYGLN--YSLMPSSG 173

```

```
QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAFS 91
DB 174 SANTAELARSISNKAIVVVTGWAPHMFA 202

RESULT 11
QH204_PSEAE
ID Q9H204_PSEAE PRELIMINARY; PRT; 285 AA.
AC Q9H204;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable glycine betaine-binding protein.
GN OrderedLocusNames=PA3236;
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H. Jr., Hancock R.E.W., Lory S., Olson M.V.;
RT "complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AE004746; AAG06624.1; -; Genomic_DNA.
DR PIR; H83241; H83241.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
DR Complete proteome.
SQ SEQUENCE 285 AA; 31080 MW; CEBALABE1F9771B9 CRC64;

Query Match 41.1%; Score 193; DB 2; Length 285;
Best Local Similarity 42.7%; Pred. No. 1.1e-11;
Matches 38; Conservative 20; Mismatches 29; Indels 2; Gaps 1;
[1]
QY 3 GFVPSYVPISSISLKGDKFKNMIGIDAGTQIVTEQALNYGLSKYEYLVPSSE 62
DB 116 GLIVPEYVSANSIADLQAKDAFGGRVVGIDAGAGVMTDKAIKQYGL--DYKLVSAG 173

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAFS 91
DB 174 SCMIATRAENKPKPVVVTGWAPHMFA 202

RESULT 12
Q4LV54_9BURK
ID Q4LV54_9BURK PRELIMINARY; PRT; 287 AA.
AC Q4LV54;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Substrate-binding region of ABC-type glycine betaine transport system
DE precursor.
GN ORFNames=Bcen2424DRAFT 5080;
OS Burkholderia cenocepacia H12424.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=331272;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H12424;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,

RA Hammon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia
RT H12424.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H12424;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia cenocepacia
RT H12424.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAHL01000012; EAM20026.1; -; Genomic_DNA.
FT SIGNAL.
SQ SEQUENCE 287 AA; 30720 MW; 3B8B86B1EB3D9DD9 CRC64;

Query Match 41.1%; Score 193; DB 2; Length 287;
Best Local Similarity 42.7%; Pred. No. 1.1e-11;
Matches 38; Conservative 16; Mismatches 33; Indels 2; Gaps 1;
[1]
QY 3 GFVPSYVPISSISLKGDKFKNMIGIDAGTQIVTEQALNYGLSKYEYLVPSSE 62
DB 116 GLIVPEYVKAISIDDLNAEKGSGFGRVIGIDAGAGVMTDKAIKSYGLN--YTLMPSSG 173

QY 63 SVMLASLDSSIKRNEWILVPLWKPHWAFS 91
DB 174 SANTAELRSYGVANKPKPVITGWAPHMFA 202

RESULT 13
Q828F3_STRAW
ID Q828F3_STRAW PRELIMINARY; PRT; 871 AA.
AC Q828F3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative glycine betaine ABC transport system permease protein.
GN Name=OpuABC; OrderedLocusNames=SAV6716;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=11572948; DOI=10.1038/nbt820;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531 (2003).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
CC -!- FUNCTION: Part of a binding-protein-dependent transport system.
CC Probably responsible for the translocation of the substrate across
CC the membrane (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the binding-protein-dependent transport
CC system permease family.
CC EMBL; BA000030; BAC74427.1; -; Genomic_DNA.
```

```
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; F:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF00528; BPD transp_1; 1.
DR Pfam; PF04069; OpuAC; 2.
DR PROSITE; PS00928; ABC_TM1; 1.
KW Complete proteome; Transmembrane; Transport.
SQ SEQUENCE 871 AA; 94924 MW; E6D04F3253C8925C CRC64;

Query Match 40.9%; Score 192; DB 2; Length 871;
Best Local Similarity 45.5%; Pred. No. 5e-11;
Matches 40; Conservative 15; Mismatches 31; Indels 2; Gaps 2;

QY 6 VPSYV-PISSISELKGDKFKNMIGIDAGAG-TQIVTEQALNYGLSKEYELVPSSE 63
Db VPSYVVKVSLADLKGKGTGDKIGIPGTGEMNLLKTKVLPGLDKEYKVVDDGTF 758
QY 64 VMLASLSDSSIKRNEWILVPLWKPHWAFS 91
Db 759 AMLAEKRAYAKKEPVAVVLWSPHWAYS 786

RESULT 14
QAKFZ7_PSEFS
ID QAKFZ7_PSEFS PRELIMINARY; PRT; 283 AA.
AC QAKFZ7
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Glycine betaine/L-proline ABC transporter, glycine betaine/L-proline-
DE binding/permease protein.
GN Name=prox; ORFNames=PFL_1710;
OS Pseudomonas fluorescens (strain Pf-5).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=220664;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Pf-5;
RX PubMed=15980861; DOI=10.1038/nbt1110;
RA Paulsen I.T., Press C., Ravel J., Kobayashi D., Myers G.S.,
RA Mavrodi D., DeBoy R.T., Seshadri R., Ren Q., Madupu R., Dodson R.J.,
RA Durkin S., Brinkac L.M., Daugherty S.C., Sullivan S.A., Rosovitz M.,
RA Gwinn M.L., Zhou L., Nelson W.C., Weidman J., Watkins K., Tran K.,
RA Khouri H.M., Pierson E., Pierson L. III, Thomasow L., Loper J.;
RT "Complete genome sequence of the plant commensal Pseudomonas
RT fluorescens Pf-5";
RL Nat. Biotechnol. 23:873-878(2005).
DR EMBL; CP000076; AAY91005.1; -; Genomic DNA.
SQ SEQUENCE 283 AA; 30765 MW; 85963158ED83E482 CRC64;

Query Match 39.9%; Score 187.5; DB 2; Length 283;
Best Local Similarity 43.8%; Pred. No. 4.2e-11;
Matches 39; Conservative 17; Mismatches 30; Indels 3; Gaps 2;

QY 3 GFVVPSPYVPISSISELKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62
Db 114 GLIVPEYVKAISIEDLK-TDTTFKNKIVGIDAGSGVWLKTDQAIKDYGL--DYKLQASSG 170
QY 63 SVMLASLSDSSIKRNEWILVPLWKPHWAFS 91
Db 171 AAMIAELTRAEKQDSIAVTGWVPHMFA 199

RESULT 15
Q9KIF6_9LACT
ID Q9KIF6_9LACT PRELIMINARY; PRT; 573 AA.
AC Q9KIF6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE OpuABC.
CN Name=opuABC;
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20318987; PubMed=10860977; DOI=10.1073/pnas.97.13.7102;
RA van der Heide T., Poolman B.;
RT "Osmoregulated ABC-transport system of Lactococcus lactis senses water
RT stress via changes in the physical state of the membrane.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:7102-7106(2000).
DR EMBL; AF234619; AAF37879.1; -; Genomic_DNA.
DR PIR; C86806; C86806.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0015450; F:protein translocase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR InterPro; IPR002208; SecY.
DR Pfam; PF00528; BPD transp_1; 1.
DR Pfam; PF04069; OpuAC; 1.
DR PRINTS; PR00303; SECYTRNLCASE.
DR PROSITE; PS00928; ABC_TM1; 1.
SQ SEQUENCE 573 AA; 61952 MW; DFB59CEAB8462A0F CRC64;

Query Match 39.7%; Score 186.5; DB 2; Length 573;
Best Local Similarity 40.4%; Pred. No. 1.2e-10;
Matches 36; Conservative 21; Mismatches 29; Indels 3; Gaps 1;

QY 3 GFVVPSPYVPISSISELKGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSE 62
Db 406 GFVVPSPYVNVNSIEDLTNQANK---TITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSS 462
QY 63 SVMLASLSDSSIKRNEWILVPLWKPHWAFS 91
Db 463 GAMTVLGEALQKHKDIVITGMSPHMFM 491

Search completed: January 24, 2006, 19:56:43
Job time : 22.5539 secs
```


SQ Sequence 290 AA;
 Query Match 100.0%; Score 1527; DB 2; Length 290;
 Best Local Similarity 100.0%; Pred. No. 5e-137;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MYKFLFFIIIFMFLSCDEKSKSNLKVIGYVNWGGTAATNVLKVFEKMGYNNAEIPS 60
 1 MYKFLFFIIIFMFLSCDEKSKSNLKVIGYVNWGGTAATNVLKVFEKMGYNNAEIPS 60
 61 VTTSTMYQYLAGSKIDGTSSWVPTADKFFYEKLTKTFVDLGNAYEGTIQGFVPSYVPI 120
 61 VTTSTMYQYLAGSKIDGTSSWVPTADKFFYEKLTKTFVDLGNAYEGTIQGFVPSYVPI 120
 121 SSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSVWLASLDSS 180
 121 SSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSVWLASLDSS 180
 181 IKRNEWILVPLWKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVFDH 240
 181 IKRNEWILVPLWKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVFDH 240
 241 FYWSDLLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKTWVPEKYKTLFD 290
 241 FYWSDLLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKTWVPEKYKTLFD 290
 RESULT 2
 ID ADU98749
 AC ADU98749
 XX ADU98749;
 24-FEB-2005 (first entry)
 DE Borrelia burgdorferi antigenic polypeptide seqid 8.
 XX
 KW antibacterial; vaccine; immune stimulation; immunity; antigen;
 KW DNA library.
 XX Borrelia burgdorferi.
 OS
 XX WO2004103269-A2.
 PN
 PD 02-DEC-2004.
 PF 17-OCT-2003; 2003WO-US033056.
 XX
 PR 18-OCT-2002; 2002US-0419401P.
 XX
 PA (MACR-) MACROGENICS INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Sykes KF, Hale KS, Johnston SA;
 XX
 DR WPI; 2004-834155/82.
 DR N-PSDB; ADU98749.
 XX
 PT Immunizing a subject against Borrelia burgdorferi infection comprises
 PT providing to the subject at least one Borrelia antigen or its fragment.
 XX
 PS Claim 27; SEQ ID NO 8; 121pp; English.
 XX
 CC The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;

CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 290 AA;
 Query Match 100.0%; Score 1527; DB 8; Length 290;
 Best Local Similarity 100.0%; Pred. No. 5e-137;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 MYKFLFFIIIFMFLSCDEKSKSNLKVIGYVNWGGTAATNVLKVFEKMGYNNAEIPS 60
 1 MYKFLFFIIIFMFLSCDEKSKSNLKVIGYVNWGGTAATNVLKVFEKMGYNNAEIPS 60
 61 VTTSTMYQYLAGSKIDGTSSWVPTADKFFYEKLTKTFVDLGNAYEGTIQGFVPSYVPI 120
 61 VTTSTMYQYLAGSKIDGTSSWVPTADKFFYEKLTKTFVDLGNAYEGTIQGFVPSYVPI 120
 121 SSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSVWLASLDSS 180
 121 SSISELKGKDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSVWLASLDSS 180
 181 IKRNEWILVPLWKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVFDH 240
 181 IKRNEWILVPLWKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVFDH 240
 241 FYWSDLLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKTWVPEKYKTLFD 290
 241 FYWSDLLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKTWVPEKYKTLFD 290
 RESULT 3
 AAY19955
 ID AAY19955 standard; protein; 275 AA.
 XX
 AC AAY19955;
 19-JUL-1999 (first entry)
 DE B. burgdorferi antigenic protein, t810.aa.
 XX
 KW Antigenic protein; vaccine; Lyme disease; infection; detection.
 XX
 OS Borrelia burgdorferi.
 XX
 PN WO9859071-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US012719.
 XX
 PR 20-JUN-1997; 97US-0050359P.
 PR 22-JUL-1997; 97US-0053344P.
 PR 22-JUL-1997; 97US-0053377P.
 PR 03-SEP-1997; 97US-0057483P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMMUNE INC.
 XX
 PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 DR WPI; 1999-189980/16.
 DR N-PSDB; AAX61652.
 XX
 PT New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of diseases caused
 PT by Borrelia, particularly Lyme disease.
 XX
 PS Claim 12; Page 134; 275pp; English.

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 64.8074 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-8
Perfect score: 1527
Sequence: 1 MYKLFPIIFIMFLSCDEKK.....EKNKEIVKTWPEKYKTLFD 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq 21:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1527	100.0	290	2 AAY19954	Aay19954 B. burgdo
2	1527	100.0	290	8 ADU98749	Adu98749 Borrelia
3	1451	95.0	275	2 AAY19955	Aay19955 B. burgdo
4	493	32.3	577	7 ADC95590	Adc95590 E. faeciu
5	468	30.6	91	8 ADU98747	Adu98747 Borrelia
6	459.5	30.1	309	9 AEM96406	Aem96406 M. xanthu
7	459	30.1	569	4 AU34910	Au34910 Enterococ
8	455	29.8	571	7 ADH87483	Adh87483 Enterococ
9	451	29.5	573	5 ABB54794	Abb54794 Lactococc
10	442.5	29.0	575	5 ABP27731	Abp27731 Streptoco
11	439	28.7	575	5 ABP27730	Abp27730 Streptoco
12	439	28.7	575	5 ABP29826	Abp29826 Streptoco
13	439	28.7	575	8 ADV88795	Adv88795 Streptoco
14	439	28.7	575	8 ADV80048	Adv80048 Streptoco
15	439	28.7	575	8 ADW82185	Adw82185 Streptoco
16	438	28.7	384	7 ABO72983	Abo72983 Pseudomon
17	433.5	28.4	320	6 ADB07852	Adb07852 Alloioioc
18	416.5	27.3	334	6 ADB07856	Adb07856 Alloioioc
19	416.5	27.3	352	6 ADB07858	Adb07858 Alloioioc
20	408	26.7	302	6 ADB11550	Adb11550 Alloioioc
21	406	26.6	303	6 ADB07978	Adb07978 Alloioioc
22	406	26.6	326	6 ADB07980	Adb07980 Alloioioc
23	338.5	22.2	273	6 ADB07854	Adb07854 Alloioioc
24	315	20.6	349	4 AAU46688	Aau46688 Propionib

25	315	20.6	349	6 ABM43207	Abm43207 Propionib
26	294	19.3	417	4 AAU40421	Aau40421 Propionib
27	294	19.3	417	6 ABM36940	Abm36940 Propionib
28	279	18.3	300	5 ABB48738	Abb48738 Listeria
29	269	17.6	356	7 ABO68411	Abo68411 Pseudomon
30	262	17.2	296	8 ADG32088	Adg32088 Mutant B
31	246.5	16.1	427	7 ABO73023	Abo73023 Pseudomon
32	235	15.4	312	4 AAU33717	Aau33717 Pseudomon
33	235	15.4	312	6 ABU15624	Abu15624 Protein e
34	235	15.4	318	7 ABO68427	Abo68427 Pseudomon
35	224	14.7	380	6 ABO21819	Abo21819 Protein e
36	189	12.4	314	6 ABU39563	Abu39563 Protein e
37	175	11.5	333	6 ABM69068	Abm69068 Photorhab
38	166	10.9	361	7 ABO84191	Abo84191 Pseudomon
39	160.5	10.5	352	7 ABO84186	Abo84186 Pseudomon
40	144.5	9.5	338	7 ABO63655	Abo63655 Klebsiell
41	118	7.7	158	5 ABP40562	Abp40562 Staphyloc
42	118	7.7	158	8 ADS06144	Ads06144 Staphyloc
43	118	7.7	352	7 ADF05311	Adf05311 Bacterial
44	113	7.4	165	5 ABP40486	Abp40486 Staphyloc
45	113	7.4	165	8 ADS06036	Ads06036 Staphyloc

ALIGNMENTS

RESULT 1
AAY19954
ID AAY19954 standard; protein; 290 AA.
XX
AC AAY19954;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, f810.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
PN WO9859071-Al.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US012718.
XX
PR 20-JUN-1997; 97US-0050359P.
PR 22-JUL-1997; 97US-0053344P.
PR 22-JUL-1997; 97US-0053377P.
PR 03-SEP-1997; 97US-0057483P.
XX
(HUMA-) HUMAN GENOME SCI INC.
(MEDI-) MEDIMMUNE INC.
XX
Choi GH, Erwin AL, Hanson MS, Lathigra R;
WPI; 1999-189980/16.
DR N-P8DB; AAX61651.
XX
New isolated Borrelia burgdorferi nucleic acids - used to develop products for the diagnosis, prevention and treatment of diseases caused by Borrelia, particularly Lyme disease.
XX
Claim 12; Page 134; 275pp; English.
XX
This sequence represents a Borrelia burgdorferi (Bb) protein of the invention, which is suitable for use in a vaccine. The Bb polypeptides can be used in vaccines for eliciting protective antibodies to members of the Borrelia genus, particularly for the use against Lyme disease in humans and animals. They can be used for preventing or attenuating an infection caused by a member of the Borrelia genus. The products can also be used for detection of members of the Borrelia genus

XX This sequence represents a *Borrelia burgdorferi* (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the *Borrelia* genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the *Borrelia* genus. The products can also
 CC be used for detection of members of the *Borrelia* genus

XX Sequence 275 AA;

Query Match 95.0%; Score 1451; DB 2; Length 275;
 Best Local Similarity 100.0%; Pred. No. 8.4e-130; Indels 0; Gaps 0;
 Matches 275; Conservative 0; Mismatches 0;
 QY 16 CDEKSSKSLKSVKIGYVNWGGTAAATNLKVVFERMGYNABIFSVTTSIMYQYLASGKI 75
 Db |||||
 QY 1 CDEKSSKSLKSVKIGYVNWGGTAAATNLKVVFERMGYNABIFSVTTSIMYQYLASGKI 60
 Db |||||
 QY 76 DGTSSSWPTADKFYFEKLTKEFDLGNAYEGTIQGFVVPVPISSISELKGKDGKPKN 135
 Db |||||
 QY 61 DGTSSSWPTADKFYFEKLTKEFDLGNAYEGTIQGFVVPVPISSISELKGKDGKPKN 120
 Db |||||
 QY 136 KMIGIDAGAGTQIVTEQALNYYGLSKEYELVPSSSVMLASLDSSI KRNEWILVPLWKPH 195
 Db |||||
 QY 121 KMIGIDAGAGTQIVTEQALNYYGLSKEYELVPSSSVMLASLDSSI KRNEWILVPLWKPH 180
 Db |||||
 QY 196 WAFSRDYIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYYYVDFHFYMSDDLILPLMDKN 255
 Db |||||
 QY 181 WAFSRDYIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYYYVDFHFYMSDDLILPLMDKN 240
 Db |||||
 QY 256 DKEPKERYNAVEFVERKKEIVKTWPEKYKTLFD 290
 Db |||||
 QY 241 DKEPKERYNAVEFVERKKEIVKTWPEKYKTLFD 275
 Db |||||

RESULT 4

ADCS95590
 ID ADCS95590 standard; protein; 577 AA.

XX AC ADCS95590;

XX DT 01-JAN-2004 (first entry)

XX E. faecium protein sequence SEQ ID 5217.

XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 XX abdominal-pelvic infection.

XX OS Enterococcus faecium.

XX PN US6583275-B1.

XX PD 24-JUN-2003.

XX PF 30-JUN-1998; 98US-00107532.

XX PR 02-JUL-1997; 97US-0051571P.

XX PR 14-MAY-1998; 98US-0085598P.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX DR WPI; 2003-799836/75.

XX DR N-PSDB; ADC91936.

XX New isolated nucleic acid derived from *Enterococcus faecium* encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and
 PT treatment of a pathological condition resulting from a bacterial
 PT infection.

XX Example 1; SEQ ID NO 5217; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from
 CC Enterococcus faecium encoding an *Enterococcus faecium* polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids are useful for diagnosing pathological conditions
 CC resulting from *E. faecium* bacterial infection (e.g. urinary tract
 CC infection), bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of *Candida albicans* -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating *Enterococcus faecium* infections. The present sequence represents
 CC one if the disclosed *E. faecium* proteins.

XX Sequence 577 AA;

Query Match 32.3%; Score 493; DB 7; Length 577;
 Best Local Similarity 37.0%; Pred. No. 6.9e-38;
 Matches 97; Conservative 65; Mismatches 92; Indels 8; Gaps 4;
 QY 21 SSKNLKSVKIGYVNWGGTAAATNLKVVFERMGYNABIFSVTTSIMYQYLASGKIDGTVS 80
 Db |||||
 QY 320 SAKETKRLNSVENVDETVASTNVVGEVLKQMGYDVTMTPLDNTSMKSVNGESDAMVS 379
 Db |||||
 QY 81 SWPTADKFYFEKLTKEFDLGNAYEGTIQGFVVPVPISSISELKGKDGKPKMIGI 140
 Db |||||
 QY 380 AWPKTHGSGYQAKYQKQVEDLGNLTGAKVLAVPAYMDVNSIDELTDQAGK---KIIGI 436
 Db |||||
 QY 141 DAGAGTQIVTEQALNYYGLSKEYELVPSSSVMLASLDSSI KRNEWILVPLWKPHAFSR 200
 Db |||||
 QY 437 EPAGVYTAENTIKYDNLKDKWKVETSSSGANTVALQAIKKEPIVVTGTPHWWFAK 496
 Db |||||
 QY 201 YDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYYYVDFHFYMSDDLILPLMD- KNDKE 258
 Db |||||
 QY 497 YDLKYLEDPENGMGSEQIHTWVRKGLKEDQPEAYKVLDFHWSKDKMEKVMLEINNGKD 556
 Db |||||
 QY 259 PKERYNAVEFVERKKEIVKTW 280
 Db |||||
 QY 557 P---QQAADWKIKENQELVESW 575
 Db |||||

RESULT 5

ADU98747

ID ADU98747 standard; protein; 91 AA.

XX AC ADU98747;

XX DT 24-FEB-2005 (first entry)

XX *Borrelia burgdorferi* antigenic polypeptide seqid 6.

XX antibacterial; vaccine; immune stimulation; immunity; antigen;
 XX DNA library.

XX OS *Borrelia burgdorferi*.

XX PN WO2004103269-A2.

XX PD 02-DEC-2004.

XX PF 17-OCT-2003; 2003WO-US033056.

XX PR 18-OCT-2002; 2002US-0419401P.

XX PA (MACR-) MACROGENICS INC.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Sykes KF, Hale KS, Johnston SA;
XX WPI: 2004-834155/82.
DR N-PSDB; ADU98746.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises
PT providing to the subject at least one Borrelia antigen or its fragment.
XX
XX Claim 27; SEQ ID NO 6; 121pp; English.
XX
CC The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide;
CC preparing a vaccine; vaccinating a subject; treating a subject infected
CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
CC ; and detecting antibodies to a herpesvirus. The method is useful in
CC immunizing a subject against Borrelia burgdorferi infection. This is the
CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
CC useful in immunizing a subject against infection by a member of the
CC Borrelia species.
XX
SQ Sequence 91 AA;
Query Match 30.6%; Score 468; DB 8; Length 91;
Best Local Similarity 98.9%; Pred. No. 1.3e-36;
Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 109 IQGFVVPSVYPISSISELKGDKFKFNKMGIDAGAGTQIVTEQALNYGLSKYELVPS 168
Db 1 LQGFVVPSVYPISSISELKGDKFKFNKMGIDAGAGTQIVTEQALNYGLSKYELVPS 60
QY 169 SESVMLASLDSSIKRNEWILVPLKPHWAFS 199
Db 61 SESVMLASLDSSIKRNEWILVPLKPHWAFS 91
RESULT 6
ABM96406
ID ABM96406 standard; protein; 309 AA.
XX
AC ABM96406;
DT 02-JUN-2005 (first entry)
XX
DE M. xanthus protein sequence, seq id 15605.
XX
KW Transgenic plant; DNA replication; gene regulation; gene expression.
XX
OS Myxococcus xanthus.
XX
PN US6833447-B1.
XX
PD 21-DEC-2004.
XX
PF 10-JUL-2001; 2001US-00902540.
XX
PR 10-JUL-2000; 2000US-0217883P.
XX
XX (MONS) MONSANTO TECHNOLOGY LLC.
XX
XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX WPI: 2005-028716/03.
XX
XX New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,

PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
XX Example 2; SEQ ID NO 15605; 25pp; English.
XX
CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 9692-16825 represent
CC a group of 7134 Myxococcus xanthus proteins and peptides. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
XX
SQ Sequence 309 AA;
Query Match 30.1%; Score 459.5; DB 9; Length 309;
Best Local Similarity 37.7%; Pred. No. 4.6e-35;
Matches 98; Conservative 50; Mismatches 103; Indels 9; Gaps 4;
QY 28 VKIGYVNWGGTAATNLKVVFE-RMGYNABIFSVTTSIMYOYLAGKIDGTVSSWVPTA 86
Db 54 VRLVYVNWAEGVAMTHLVQAILEDPMGEVKTWADVAPVFAASLANGADAFDGLWLPVT 113
QY 87 DKFYEKLTKEFDLGANYEGTIOGFVVPYVPISSISELKGDKFKFNKMGIDAGAGT 146
Db 114 HQSYMERFQGVKVDLGTNYEDARIGLVVPAOLDITSIEQLNGKAKDLNQSVIGIDSGAGI 173
QY 147 QIVTEQALNYGLSKYELVPSSESVMLASLDSSIKRNEWILVPLKPHWAFSRYDIKFL 206
Db 174 MTTTEKAIAEYKL--DLKLVSSGPMATLAKKRPVVVTGKPKHWKFAEWDLKL 231
QY 207 DDPDLIMGISVHTLVRLGLENDDFDAYYVDFHFWSDDLLILPLMD---KNDKEPGKEY 263
Db 232 DDPKGVYGAKESTHTLTVGLEKDLPDVATLLRNFKLDDQQLGSLMGAIAEAEDAPEKAT 291
QY 264 RNAVEFVEKNKEIVKTVYPE 283
Db 292 R---EWVKKNQALVDGWIPK 308
RESULT 7
AAU34910
ID AAU34910 standard; protein; 569 AA.
XX
AC AAU34910;
DT 14-FEB-2002 (first entry)
XX
DE Enterococcus faecalis cellular proliferation protein #197.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI: 2001-611495/70.
 DR N-PSDB; AAS52769.
 XX
 XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 10503; 511pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
 CC coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 569 AA;

Query Match 30.1%; Score 459; DB 4; Length 569;
 Best Local Similarity 35.5%; Pred. No. 1.2e-34;
 Matches 93; Conservative 59; Mismatches 102; Indels 8; Gaps 4;
 QY 21 SSKNLKSVKIGVNWGGTAATNLKVVFPEKMGYNAEIFSVTTSIMYQYLAGSKIDGTVS 80
 DB 312 TTNDKQISLSVEMDTEVASTHVAEVLKDMGYDKTTPLDNALMWSVAKGETDAMVG 371
 QY 81 SWVPADKFFYEKLTKEFDLGANYEGTIGQFVPSYVPISSISLKGKDKFKNMIGI 140
 DB 372 AWPFGTHAEQYKQYKDKLDDLGENLKGAKLGIVPSYMDVDSIEDLSQAGK---KITGI 428
 QY 141 DAGAGTQIVTEQALNYGLSKYEYLVPSSESVMLASLSSIKRNEWILVPLKPHWAFSR 200
 DB 429 EPGAGVAAAEKTKAEYPNLKDWSVETSSSGAMTVAGQAIKNEDIVITGWSPHWMAK 488
 QY 201 YDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVDFHFW-SDDLILPLMDKNDKEP 259
 DB 489 YDLKYLADPKGTMGGEAHTHARQGLKEDQPEAYKVLDFNFWHTTKDMESVMLEINE--- 545
 QY 260 GKEYRNAV-EFEKNEKEIVKTW 280
 DB 546 GKDPQEAARDWVDSHKQVAEW 567

RESULT 8
 ADH87483
 ID ADH87483 standard; protein; 571 AA.
 XX
 AC ADH87483;
 XX
 DT 22-APR-2004 (first entry)
 XX
 DE Enterococcus faecalis polypeptide #1963.
 XX
 KW Enterococcus faecalis infection; transcription regulatory element;
 KW antibacterial.
 XX
 OS Enterococcus faecalis.
 XX

PN US6617156-B1.
 XX
 PD 09-SEP-2003.
 XX
 PF 13-AUG-1998; 98US-00134000.
 XX
 PR 15-AUG-1997; 97US-0055778P.
 XX
 PA (DOUC/) DOUCETTE-STAMM L A.
 PA (BUSH/) BUSH D.
 XX
 PI Doucette-Stamm LA, Bush D;
 XX
 DR WPI: 2003-895394/82.
 DR N-PSDB; ADH84078.
 XX
 PT New nucleic acid comprising a sequence encoding an *Enterococcus faecalis*
 PT polypeptide, useful for preparing a composition for diagnosing or
 PT treating *E. faecalis* infection.
 XX
 PS Disclosure; SEQ ID NO 5368; 193pp; English.
 XX
 CC The invention relates to *Enterococcus faecalis* polynucleotides and
 CC polypeptides. The invention also relates to a recombinant expression
 CC vector comprising a polynucleotide operably linked to a transcription
 CC regulatory element, a cell comprising a recombinant vector, a method for
 CC producing an *E. faecalis* polypeptide, an isolated nucleic acid comprising
 CC a sequence not given in the specification, a recombinant vector
 CC comprising the nucleic acid and a cell comprising the recombinant vector.
 CC The polynucleotides can be used to detect the presence of *E. faecalis* in
 CC a sample. The sequences are useful for preparing a composition for
 CC diagnosing or treating *Enterococcus faecalis* infection. This sequence
 CC represents an *E. faecalis* polypeptide of the invention.
 XX
 SQ Sequence 571 AA;

Query Match 29.8%; Score 455; DB 7; Length 571;
 Best Local Similarity 35.1%; Pred. No. 2.9e-34;
 Matches 92; Conservative 60; Mismatches 102; Indels 8; Gaps 4;
 QY 21 SSKNLKSVKIGVNWGGTAATNLKVVFPEKMGYNAEIFSVTTSIMYQYLAGSKIDGTVS 80
 DB 314 TTNDKQISLSVEMDTEVASTHVAEVLKDMGYDKTTPLDNALMWSVAKGETDAMVG 373
 QY 81 SWVPADKFFYEKLTKEFDLGANYEGTIGQFVPSYVPISSISLKGKDKFKNMIGI 140
 DB 374 AWPFGTHAEQYKQYKDKLDDLGENLKGAKLGIVPSYMDVDSIEDLSQAGK---KITGI 430
 QY 141 DAGAGTQIVTEQALNYGLSKYEYLVPSSESVMLASLSSIKRNEWILVPLKPHWAFSR 200
 DB 431 EPGAGVAAAEKTKAEYPNLKDWSVETSSSGAMTVAGQAIKNEDIVITGWSPHWMAK 490
 QY 201 YDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVDFHFW-SDDLILPLMDKNDKEP 259
 DB 491 YDLKYLADPKGTMGGEAHTHARQGLKEDQPEAYKVLDFNFWHTTKDMESVMLEINE--- 547
 QY 260 GKEYRNAV-EFEKNEKEIVKTW 280
 DB 548 GKDPQEAARDWVDSHKQVAEW 569

RESULT 9
 ABB54794
 ID ABB54794 standard; protein; 573 AA.
 XX
 AC ABB54794;
 XX
 DT 29-AUG-2003 (revised)
 DT 16-MAY-2002 (first entry)
 XX
 XX Lactococcus lactis protein busAB.
 DE Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 KW

XX Lactococcus lactis; IL11403.
 XX FR2807446-A1.
 XX 12-OCT-2001.
 XX 11-APR-2000; 2000FR-00004630.
 XX 11-APR-2000; 2000FR-00004630.
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 XX WPI; 2002-043418/06.
 XX New nucleotide sequence useful in the identification or Lactococcus
 XX lactis and related species.
 XX Claim 6; SEQ ID NO 1496; 2504pp; French.
 XX The present invention is related to a Lactococcus lactis nucleotide
 XX sequence (ABA90521) and related proteins (AB53300-AB55621). The nucleic
 XX acid sequence is useful in the detection and/or amplification of nucleic
 XX acid sequence, particularly to identify Lactococcus lactis or related
 XX species. The proteins of the invention are useful for the biosynthesis or
 XX biodegradation of a composition of interest. The invention helps research
 XX in lactic bacteria, particularly useful in the production of yogurt and
 XX cheese. Note: The sequence data for this patent is based on equivalent
 XX patent WO200177334 (published 18-OCT-2001) which is available in
 XX electronic format directly from WIPO at
 XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 XX standardise OS field)
 XX Sequence 573 AA;
 Query Match 29.5%; Score 451; DB 5; Length 573;
 Best Local Similarity 34.9%; Pred. No. 7e-34;
 Matches 99; Conservative 64; Mismatches 99; Indels 22; Gaps 7;
 QY 6 LFFIIFMFLSCDEKSSKNLKVIGYVNWGGTAATNVLKVVFERKMGYNABIFSVTTSI 65
 Db 302 LALIIGAFSGMSFGKTASD-KKVDLVYMNWDSEVASINVLTAQKEHGFVKTTALDNAV 360
 QY 66 MTQYLASGKIDGTVSWVFTADKFFYEKLTKEFVD-LGANYEGTIGQFVVPVPISSIS 124
 Db 361 AWQTVANGQADGWSAWLENTHKTQMKY-GKSVDLLGPNLKGAKVGFVVPVYMNVSIE 419
 QY 125 ELKGGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSSEVNLASLDSIKRN 184
 Db 420 DLTNQANK---TITGLEPGAGVMAASEKTLSYDNLKQWLVPSSESGMTVALGSAIKQH 476
 QY 185 EWILVPLKPHWAFGRYDIKFLDDPDLMINGGIESVHTLVRLGLENDDFDAYVVFHFYV- 243
 Db 477 KDIVITGWSPHMKNKYDLKYLADPKGTGTSENINTIVRGKLNKPEAYKVLQKFWNT 536
 QY 244 SDDLILPLMD-KNDKEPGKEYNVAFVEKKNKEIVKTWVPEKYK 286
 Db 537 TKDMEAVMLDIQNGKTP-----BEAAKNWIKDHQK 566
 RESULT 10
 ABP27731
 ID ABP27731 standard; protein; 575 AA.
 XX AC
 XX ABP27731;
 XX 02-JUL-2002 (first entry)
 DT Streptococcus polypeptide SEQ ID NO 4638.
 DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX Streptococcus pyogenes.
 OS WO200234771-A2.
 PN 02-MAY-2002.
 PD 29-OCT-2001; 2001WO-GB004789.
 PF 27-OCT-2000; 2000GB-00026333.
 PR 24-NOV-2000; 2000GB-00028727.
 PR 07-MAR-2001; 2001GB-00005640.
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
 PI Tettelin H;
 PI WPI; 2002-352536/38.
 DR N-PSDB; ABN68362.
 XX New Streptococcus protein for the treatment or prevention of infection or
 XX disease caused by Streptococcus bacteria, such as meningitis, and for
 XX detecting a compound that binds to the protein.
 PS Claim 1; Page 3626-3627; 4525pp; English.
 XX The invention relates to a protein (ABP25413-ABP30895) from group B
 XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 XX (Streptococcus pyogenes), comprising one of 5493 sequences (S1), given in
 XX the specification. The proteins have antibacterial and antiinflammatory
 XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 XX antibodies that bind (I) are used in the manufacture of medicaments for
 XX the treatment or prevention of infection or disease caused by
 XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 XX Nucleic acids encoding (I) are used to detect Streptococcus in a
 XX biological sample. (I) is used to determine whether a compound binds to
 XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 XX used as a vaccine or diagnostic composition. The disease caused by
 XX Streptococcus that is prevented or treated may be meningitis. Nucleic
 XX acid encoding (I) may be used to recombinantly produce (I) and may be
 XX used in gene therapy. Antibodies to (I) are used for affinity
 XX chromatography, immunossays, and distinguishing/identifying
 XX Streptococcus proteins
 XX Sequence 575 AA;
 Query Match 29.0%; Score 442.5; DB 5; Length 575;
 Best Local Similarity 35.8%; Pred. No. 4.6e-33;
 Matches 92; Conservative 56; Mismatches 102; Indels 7; Gaps 5;
 QY 26 KSVKIGYVNWGGTAATNVLKVVFERKMGYNABIFSVTTSIMYQYLASGKIDGTVSSWVPT 85
 Db 322 ETVNIAVYQWQSEVASTHVAEVLKNEGHVTLTFLDNRVMMQTVANGNADFSTGAWLPV 381
 QY 86 ADKFYIEKLTKEFVDLGANYEGTIGQFVVPVSVV-PISSISELKGKDKFKNMIGIDAGA 144
 Db 382 THGQYQYKYSKLDLDPNLKGTGLGAVPKYMTVDVNSIEDLSKQADQ---KITGEPGA 438
 QY 145 GTQIVTEQALNYGLSKEYELVPSSSEVNLASLDSIKRNKEIVLPLKPHWAFGRYDIK 204
 Db 439 GIMAAQAQTLKEYHNLSSWELVAASTGAMTTSLDQAIKKOPIVVTAWSPHWMFAKYDLK 498
 QY 205 FLDDPDLMINGGIESVHTLVRLGLENDDFDAYVVFHFYVSD-DLILPLMDKNDKEPGKE 263
 Db 499 YLKDPEKTFGTENINTIARKGLKELPNVYKIIDKFHTQKMEAVMLDIN-KGMSPB- 556
 QY 264 RNAVEFVEKKNKEIVKTW 280
 Db 557 AAARKWVEANKSVSSW 573

Db 366 TVANGNADFTTSAMLKPTHGQYFNKYKNSLDDLGPHVENVKIGLVVFKMNVNSTEELSN 425

QY 129 KGDFKFKMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSESVNMLASDSSIKENWIL 188

Db 426 QADK---QITGIEPGAGIMKSAQSLKDYPNLSSWKLLSASTGAMTTTLGKAIKNKQV 482

QY 189 VPLWKPHWAFSRDYIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVYVDFHFWSDDLI 248

Db 483 ITGSPHWFYKDYLDKLPKPKSGFSGEHIINTIARKNLKKDMPKVYKIIDFKFKWKEDM 542

QY 249 LPLMDKNDK--EPGKEYRNAVFEVKEKKEIVKWT 280

Db 543 ESIMLMDKMGEPAKA---AQKWKIKNHRKEVSEW 573

RESULT 12

ID ABP29826 standard; protein; 575 AA.

XX ABP29826;

XX 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 8828.

DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB004789.

XX 27-OCT-2000; 2000GB-00026333.

PR 24-NOV-2000; 2000GB-00028727.

PR 07-MAR-2001; 2001GB-00005640.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;

PI Tettelin H;

XX WPI; 2002-352536/38.

DR N-PSDB; ABN68361.

XX New Streptococcus protein for the treatment or prevention of infection or disease caused by Streptococcus bacteria, such as meningitis, and for detecting a compound that binds to the protein.

XX Claim 1; Page 3626; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins

XX Query Match 28.7%; Score 439; DB 5; Length 575;

XX Best Local Similarity 33.9%; Pred. No. 9.8e-33;

XX Matches 93; Conservative 55; Mismatches 114; Indels 12; Gaps 4;

QY 9 IIFMFLSCDEKSSKNLKVGIYVNWGGETAATNLVKVFFRMGYNABIFSVTTSIMVQ 68

Db 310 IVVNTSGNEARGKQK---VKIAYQWDSEVASTNVIAEVLKSGYDVELTPLDNVAVMQ 365

QY 69 YIASKGIDGTVSSWPTADKFFYEKLTFRVDLGANVETIQGVFVVPSTVPISSISLKG 128

CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
 CC synthesis of amino acids, cell membranes, intermediate (central)
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
 CC regulatory functions, replication, transcription, translation, protein
 CC transport, adaptation to atypical conditions, sensitivity to medicines
 CC and/or analogues, functions related to transposons, biosynthesis of
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and
 CC cellular machinery. (I) are useful for the detection and/or amplification
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
 CC useful for treatment of a bacterial S. agalactiae infection. The complete
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
 CC present patent is an equivalent for the basic patent FR2824074A1, which
 CC contains only 2344 sequences.

XX Sequence 575 AA;

Query Match 28.7%; Score 439; DB 8; Length 575;
 Best Local Similarity 33.9%; Pred. No. 9.8e-33;
 Matches 93; Conservative 55; Mismatches 114; Indels 12; Gaps 4;

QY 9 IIFMFLSCDEKSSKNLKSIVKIGYVNWGGTAAATNVLKVVFEKMGYNABIFSVTTSIMYQ 68
 DB 310 IVVNTSGNEAKGQK---VKIAYVQWDSEVASTNVIAEVLKSKGYDVELTFLDNVAMVQ 365
 QY 69 YLASKGIDGTVSSWVPTADKFEYKELKTFVDLGANYECTIOGVVPSVYVPISSISELKG 128
 DB 366 TVANGNADFTTSAMLPTKTHGQYFNKYKNSLDLGHVENVKIGLVVPKYMNVNSEESEN 425
 QY 129 KGDGKFNKMGIDAGAGTQIVTEQALNYYGLSKYELVPSSESVMLASLDSSIKRNEWTL 188
 DB 426 QADK---QITGIEPGAGIMKSAKQSLKDYPNLSSWKLSSASTGAMTTTLGKAIKNDQVV 482
 QY 189 VPLKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRGLGENDDFDAYVYDFHFYSDDLI 248
 DB 483 ITGWSPHWMAFYDKYLKDPKPSFGGSEHINTIARKNLKDKMPKVKYKIIDKFKWTKEDM 542
 QY 249 LPLMDKNDK--EPGKEYRNAVEFVEKNKEIVKWTW 280
 DB 543 ESIMLMDKMGEPAKA---AQKWKHKKEVSEW 573

RESULT 15

ADV82185

ID ADV82185 standard; protein; 575 AA.

XX ADV82185;

DT 24-FEB-2005 (first entry)

XX Streptococcus agalactiae protein, SEQ ID 3326.

XX Antibacterial; vaccine; bacterial infection.

XX Streptococcus agalactiae.

XX WO200292818-A2.

XX 21-NOV-2002.

XX 26-APR-2002; 2002WO-IB0003059.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;

PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;

XX WPI; 2004-101891/11.

XX

Genomic nucleotide sequences encoding polypeptides of Streptococcus
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
 PT and identification of therapeutic targets.

PS Claim 6; SEQ ID NO 3326; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
 CC synthesis of amino acids, cell membranes, intermediate (central)
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
 CC regulatory functions, replication, transcription, translation, protein
 CC transport, adaptation to atypical conditions, sensitivity to medicines
 CC and/or analogues, functions related to transposons, biosynthesis of
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and
 CC cellular machinery. (I) are useful for the detection and/or amplification
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
 CC useful for treatment of a bacterial S. agalactiae infection. The complete
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
 CC present patent is an equivalent for the basic patent FR2824074A1, which
 CC contains only 2344 sequences.

XX Sequence 575 AA;

Query Match 28.7%; Score 439; DB 8; Length 575;
 Best Local Similarity 33.9%; Pred. No. 9.8e-33;
 Matches 93; Conservative 55; Mismatches 114; Indels 12; Gaps 4;

QY 9 IIFMFLSCDEKSSKNLKSIVKIGYVNWGGTAAATNVLKVVFEKMGYNABIFSVTTSIMYQ 68
 DB 310 IVVNTSGNEAKGQK---VKIAYVQWDSEVASTNVIAEVLKSKGYDVELTFLDNVAMVQ 365
 QY 69 YLASKGIDGTVSSWVPTADKFEYKELKTFVDLGANYECTIOGVVPSVYVPISSISELKG 128
 DB 366 TVANGNADFTTSAMLPTKTHGQYFNKYKNSLDLGHVENVKIGLVVPKYMNVNSEESEN 425
 QY 129 KGDGKFNKMGIDAGAGTQIVTEQALNYYGLSKYELVPSSESVMLASLDSSIKRNEWTL 188
 DB 426 QADK---QITGIEPGAGIMKSAKQSLKDYPNLSSWKLSSASTGAMTTTLGKAIKNDQVV 482
 QY 189 VPLKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRGLGENDDFDAYVYDFHFYSDDLI 248
 DB 483 ITGWSPHWMAFYDKYLKDPKPSFGGSEHINTIARKNLKDKMPKVKYKIIDKFKWTKEDM 542
 QY 249 LPLMDKNDK--EPGKEYRNAVEFVEKNKEIVKWTW 280
 DB 543 ESIMLMDKMGEPAKA---AQKWKHKKEVSEW 573

Search completed: January 24, 2006, 19:44:29

Job time : 65.8074 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 11.0018 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-8
Perfect score: 1527
Sequence: 1 MYKLFPLFIIPMFLSCDEKK.....EKNKEIVKTVWPEKYKTLFD 290

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	290	2 H70117	glycine betaine, L
2	451	29.5	573	2 C86806	hypothetical prote
3	438	28.7	285	2 H83241	probable glycine b
4	360	23.6	287	2 A33578	glycine betaine/l-
5	316.5	20.7	292	2 A13042	hypothetical prote
6	316.5	20.7	304	2 C98243	hypothetical prote
7	279	18.3	300	2 A11201	glycine betaine AB
8	278	18.2	300	2 A11559	glycine betaine AB
9	269	17.6	312	2 C82974	hypothetical prote
10	252	16.5	293	2 I40537	glycine betaine AB
11	246.5	16.1	307	2 D83642	hypothetical prote
12	240	15.7	312	2 D95891	probable glycine-b
13	235	15.4	312	2 H82972	hypothetical prote
14	226.5	14.8	317	2 AC3128	conserved hypothet
15	226.5	14.8	363	2 F98159	hypothetical prote
16	216	14.1	326	2 AC3307	glycine betaine/l-
17	206.5	13.5	308	2 AE2829	hypothetical prote
18	206.5	13.5	337	2 C97607	hypothetical prote
19	172	11.3	317	2 AH2856	hypothetical prote
20	159.5	10.4	340	2 H83007	hypothetical prote
21	159	10.4	322	2 E83009	probable binding p
22	146	9.6	265	2 E97633	hypothetical prote
23	141.5	9.3	333	2 G95380	probable periplasm
24	138	9.0	333	2 D95968	probable amino aci
25	137.5	9.0	331	2 AC0842	glycine betaine-bi
26	136	8.9	334	2 AC0322	glycine betaine-bi
27	133.5	8.7	426	2 A13053	hypothetical prote
28	133.5	8.7	426	2 D98232	hypothetical prote
29	132	8.6	330	1 BLRCGP	glycine betaine/pr

30	132	8.6	330	2 F91071	glycine betaine/pr
31	131.5	8.6	336	2 AG2686	ABC transporter, s
32	131.5	8.6	336	2 D97468	hypothetical prote
33	128	8.4	330	2 A85916	hypothetical prote
34	121	7.9	345	2 AG3186	hypothetical prote
35	112.5	7.4	261	2 AD2747	hypothetical prote
36	112.5	7.4	261	2 C97528	antigenic protein
37	107.5	7.0	345	2 AD3024	hypothetical prote
38	107.5	7.0	359	2 E98260	periplasmic bindin
39	104.5	6.8	319	2 AC3523	aliphatic sulfonat
40	102.5	6.7	210	2 I40272	outer surface prot
41	102	6.7	271	2 A81292	amino acid ABC tra
42	100.5	6.6	570	2 G88098	protein F18A12.3 [
43	100.5	6.6	663	2 S40308	p100 protein - Lym
44	100	6.5	271	2 A10532	probable lipoprote
45	99.5	6.5	205	2 I40105	outer surface prot

ALIGNMENTS

RESULT 1

H70117

glycine betaine, L-proline ABC transporter, glycine/betaine/L-proline-binding protein (L
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: H70117
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Karpman, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: H70117
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-290 <KLB>
A:Cross-references: UNIPROT:O51169; UNIPARC:UPI000005737B; GB:AE001125; GB:AE000783; NII
A:Experimental source: strain B31

Query Match	100.0%	Score 1527;	DB 2;	Length 290;
Best Local Similarity	100.0%	Pred. No. 1.9e-111;	Mismatches 0;	Indels 0;
Matches 290;	Conservative	0;	0;	Gaps 0;
Qy	1	MYKLFPLFIIPMFLSCDEKSKNLSVKIGVNMGGTAATNVLKVYFEKMGYNAEIPS	60	
Db	1	MYKLFPLFIIPMFLSCDEKSKNLSVKIGVNMGGTAATNVLKVYFEKMGYNAEIPS	60	
Qy	61	VTTSIMYQYLASGKIDGTVSSWVPTADKFYFEKLTKEFVDLGANYEGTIQGFVWPSYVPI	120	
Db	61	VTTSIMYQYLASGKIDGTVSSWVPTADKFYFEKLTKEFVDLGANYEGTIQGFVWPSYVPI	120	
Qy	121	SSISELKGKDKFKNMIGIDAGAGTQIVTQALNYGLSKEYELVPSSESVMASLDSS	180	
Db	121	SSISELKGKDKFKNMIGIDAGAGTQIVTQALNYGLSKEYELVPSSESVMASLDSS	180	
Qy	181	IKRNEWILVLPWKPHWAFSRVDIKFLDDPDLIMGGIESVHTLVRLGLENDPDAVYVFDH	240	
Db	181	IKRNEWILVLPWKPHWAFSRVDIKFLDDPDLIMGGIESVHTLVRLGLENDPDAVYVFDH	240	
Qy	241	FYWSDDLILPLMDKNDKEPGKEYNNAVEFVEKNKEIVKTVWPEKYKTLFD	290	
Db	241	FYWSDDLILPLMDKNDKEPGKEYNNAVEFVEKNKEIVKTVWPEKYKTLFD	290	

RESULT 2

C86806

hypothetical protein busAB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86806
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Db 32 KDLGEQINYYTITGIDAGAGIMLATQNAIKDYHLDDNNQLQTSSTAAMTSTLQAKMDKR 91

Qy 186 WLVPLWKPWFAPSYDYTKFLDDPDLIMGGIESVHTVLRLGLENDPDPAYYVDFHFW-S 244

Db 92 PIVVTGWTPHMFTKFLDKFLDDPKRNVFGNAENIHTIVKGLKBPESAYKVLDDNFFWTA 151

Qy 245 DDLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKWTW 280

Db 152 EDMSEVMLENDGVDPDE--AAKWIKNNPEKVAKW 185

RESULT 9

C82974

hypothetical protein PA5378 [imported] - Pseudomonas aeruginosa (strain PAO1)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: C82974

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: C82974

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-312 <STO>

A:Cross-references: UNIPROT:Q9HT16; UNIPARC:UPI00000C5FCF; GB:AE004950; GB:AE004091; NID:140537

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA5378

Query Match 17.6%; Score 269; DB 2; Length 312;

Best Local Similarity 28.2%; Pred. No. 1.7e-13;

Matches 75; Conservative 57; Mismatches 120; Indels 14; Gaps 6;

Qy 27 SVKIGVNWGGTAAATNVLKVFVKMGYNABIFSVTTSIMYOYLAGSKIDGVVSSWVPTA 86

Db 29 SVRFADVGVTDITVTAVTRQVLESIGYTKVKNLSVPVYRSLLANDLDFLGNWMTM 88

Qy 87 D---RFFYEKLTKFVDLGGANYEGTITQGFVFPYSYV---PISSISELKGKDKFKNKMIGI 140

Db 89 TNDIKQYAEKGTVE--TLRANLEGAKYTLAVPQYVVDGGLRSFADIAKFSDKLGNKIYGI 146

Qy 141 DAGAGTQIVTEQAL--NYGLSKVELVPSSESVMLASLDSSIKENWILVPLWKPWFAP 198

Db 147 EFGNDGNRVAQSMIDKNAFELGK-FKLVESSBAGMLSQVQRAIRNRQWVFLGWEPHPMN 205

Qy 199 SYRIKFLDDPDLIMG--GIESVHTVLRLGLENDPDPAYYVDFHFWSDDLILPLMDKN 255

Db 206 TFPQWKYLEGGDDFFGPNYGGATITVNRKGYAQECANVGQLLKNLSFTLEMENKLMADV 265

Qy 256 DKEPGKEYRNAVEFVEKNKEIVKWTW 281

Db 266 LNNKKPBEAAKAWLKHDPQLDAWL 291

RESULT 10

I40537

glycine betaine ABC transporter (glycine betaine-binding protein) opuAC precursor - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C:Accession: I40537; F69669

R:Kempf, B.; Bremer, E.

J. Biol. Chem. 270, 16701-16713, 1995

A:Title: OpuA, an osmotically regulated binding protein-dependent transport system for glycine betaine

A:Reference number: A57322; MUID:95348093; PMID:7622480

A:Accession: I40537

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-293 <RES>

A:Cross-references: UNIPROT:P46922; UNIPARC:UPI000005FED3; EMBL:U17292; NID:9984802; PID:R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero

C.; Bron, S.; Brouillette, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadai, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror, A.; Shimizu, M.; Tamakoshi, A.; Tanaka, T.; Terpatra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.P.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69669

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-293 <KUN>

A:Cross-references: UNIPARC:UPI000005FED3; GB:Z99105; GB:AL009126; NID:92632457; PIDN:CA

A:Experimental source: strain 168

C:Genetics:

A:Gene: opuAC

Query Match 16.5%; Score 252; DB 2; Length 293;

Best Local Similarity 33.1%; Pred. No. 3.3e-12;

Matches 53; Conservative 34; Mismatches 71; Indels 2; Gaps 2;

Qy 121 SSISELKGKDKFKNKMIGIDAGAGTQIVTBOALNYGLSKVELVPSSESVMLASLDSS 180

Db 23 SENDENASAAEQVNKTIIGIDPGGIMSLTDKMKDYDLN-DWTLISASSAAMTATLKKS 81

Qy 181 IKRNEWILVPLWKPWFAPSYDYTKFLDDPDLIMGGIESVHTVLRLGLENDPDPAYYVDFH 240

Db 82 YDRKKPIIITGWTPHMFPFSLYKLYLDDPKQSGSAEIHITTRKGFSGEQPNAKLLSQ 141

Qy 241 FVWSDLLILPLMDKNDKEPKGYRNAVEFVEKNKEIVKWTW 280

Db 142 FKWTQDEMGEMIKVE-EGEPKPAKVAEYVKNHKKDQIAEW 180

RESULT 11

D83642

hypothetical protein PA0030 [imported] - Pseudomonas aeruginosa (strain PAO1)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: D83642

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83642

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <STO>

A:Cross-references: UNIPROT:Q9I7A0; UNIPARC:UPI00000C4EF4; GB:AE004442; GB:AE004091; NID:140537

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA0030

Query Match 16.1%; Score 246.5; DB 2; Length 307;

Best Local Similarity 30.4%; Pred. No. 9.5e-12;

Matches 63; Conservative 42; Mismatches 91; Indels 11; Gaps 7;

Qy 28 VKIGVNWGGTAAATNVLKV-VFEKMGYNABIFSVTTSIMYOYLAGSKIDGVVSSWVPTA 86

Db 30 VRLADPGW-SDIAVTNATAAFLLSLGQVKIDTILSVPIIYVGLRDGQVDAFLGGWMP-A 87

Qy 87 DKFYFEKL--KTKFVDLGGANYEGTITQGFVFPYSYV---PISSISELKGKDKFKNKMIGID 141

Db 88 HDYHDKFVASQVERLGRNLDGTRFTLAVPRYVWDAGVHRFPEDLAAQOQRFNRKLYGIG 147

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 17.8779 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-8
Perfect score: 1527
Sequence: 1 MYKFLFLFIIFMFLSCDEKK.....EKNKEIVKTVWPEKTKTLFD 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	290	US-09-830-230A-361	Sequence 361, App
2	1451	95.0	275	US-09-830-230A-362	Sequence 362, App
3	493	32.3	577	US-09-107-532A-5217	Sequence 5217, App
4	459.5	30.1	309	US-09-902-540-15605	Sequence 15605, A
5	455	29.8	571	US-09-134-000C-5368	Sequence 5368, App
6	438	28.7	384	US-09-252-991A-21729	Sequence 21729, A
7	269	17.6	356	US-09-252-991A-17157	Sequence 17157, A
8	246.5	16.1	427	US-09-252-991A-21769	Sequence 21769, A
9	235	15.4	318	US-09-252-991A-17173	Sequence 17173, A
10	166	10.9	361	US-09-252-991A-32937	Sequence 32937, A
11	160.5	10.5	352	US-09-252-991A-32932	Sequence 32932, A
12	144.5	9.5	338	US-09-489-039A-10172	Sequence 10172, A
13	118	7.7	158	US-09-134-001C-5407	Sequence 5407, App
14	118	7.7	352	US-09-543-681A-5596	Sequence 5596, App
15	113	7.4	165	US-09-134-001C-5331	Sequence 5331, App
16	108.5	7.1	266	US-08-961-083-14	Sequence 14, Appl
17	108.5	7.1	266	US-09-536-784-14	Sequence 14, Appl
18	108.5	7.1	266	US-09-765-271-14	Sequence 14, Appl
19	108.5	7.1	266	US-09-765-272A-14	Sequence 14, Appl
20	106.5	7.0	285	US-09-489-039A-10641	Sequence 10641, A
21	100.5	6.6	337	US-09-328-352-6980	Sequence 6980, App
22	100.5	6.6	663	US-09-196-293-5	Sequence 5, Appl
23	100.5	6.6	663	US-08-209-603B-5	Sequence 5, Appl
24	100.5	6.6	663	US-09-711-546-5	Sequence 5, Appl
25	100.5	6.6	663	US-10-289-795-5	Sequence 5, Appl
26	98.5	6.5	663	US-08-235-836C-78	Sequence 78, Appl
27	97.5	6.4	823	US-09-107-532A-5667	Sequence 5667, App

28	95	6.2	221	2	US-09-107-433-2796	Sequence 2796, App
29	95	6.2	372	2	US-09-710-279-182	Sequence 182, App
30	95	6.2	391	2	US-09-134-001C-3952	Sequence 3952, App
31	95	6.2	1751	2	US-09-136-574A-44	Sequence 44, Appl
32	94	6.2	990	2	US-10-101-464A-814	Sequence 814, App
33	93.5	6.1	773	1	US-08-966-389-4	Sequence 4, Appl
34	93.5	6.1	773	1	US-09-103-509-4	Sequence 4, Appl
35	93.5	6.1	773	1	US-09-102-644-4	Sequence 4, Appl
36	93.5	6.1	773	1	US-09-218-032-4	Sequence 4, Appl
37	93.5	6.1	1426	2	US-09-136-574A-43	Sequence 43, Appl
38	92.5	6.1	700	2	US-08-235-836C-74	Sequence 74, Appl
39	92	6.0	1881	2	US-09-233-086-3	Sequence 3, Appl
40	91.5	6.0	441	2	US-09-328-352-5114	Sequence 5114, App
41	91.5	6.0	663	2	US-08-235-836C-70	Sequence 70, Appl
42	91	6.0	339	2	US-09-107-532A-5639	Sequence 5639, App
43	90.5	5.9	287	2	US-09-107-532A-4780	Sequence 4780, App
44	90.5	5.9	317	2	US-09-252-991A-31329	Sequence 31329, A
45	90	5.9	279	2	US-09-543-681A-6534	Sequence 6534, App

ALIGNMENTS

RESULT 1

US-09-830-230A-361
; Sequence 361, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 361
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-361

Query Match	100.0%	Score 1527;	DB 2;	Length 290;
Best Local Similarity	100.0%	Pred. No. 8.6e-150;	Mismatches 0;	Indels 0; Gaps 0;
Matches 290;	Conservative 0;			
Qy	1	MYKFLFLFIIFMFLSCDEKSKSNLKS	VKVGYNVNGGTAATNVLKVF	EKNKEIVKTVWPEKTKTLFD 290
Db	1	MYKFLFLFIIFMFLSCDEKSKSNLKS	VKVGYNVNGGTAATNVLKVF	EKNKEIVKTVWPEKTKTLFD 290
Qy	61	VTTSTMYQYLASGKIDGTVSSWVPTADK	FYFEKLTGKTFVDLGANYEGTIQGFV	WPSYVPI 120
Db	61	VTTSTMYQYLASGKIDGTVSSWVPTADK	FYFEKLTGKTFVDLGANYEGTIQGFV	WPSYVPI 120
Qy	121	SSISBLKGGKDKFKNMIGIDAGAGTQV	TEQALNYGLSKYEYELVPSES	VMLASLDS 180
Db	121	SSISBLKGGKDKFKNMIGIDAGAGTQV	TEQALNYGLSKYEYELVPSES	VMLASLDS 180
Qy	181	IKRNEWILVPLWKPHWAFSRDYDIKFLD	DDPDLIMGGIESVHTLVRLGLEND	DFDAYVFDH 240
Db	181	IKRNEWILVPLWKPHWAFSRDYDIKFLD	DDPDLIMGGIESVHTLVRLGLEND	DFDAYVFDH 240
Qy	241	FWSDDLILPLMDKNDKPGKRYNAVF	EKNKEIVKTVWPEKTKTLFD	290
Db	241	FWSDDLILPLMDKNDKPGKRYNAVF	EKNKEIVKTVWPEKTKTLFD	290

Db 241 FYWSDLLPLMDKNDKEGKEYRNAVEFVEKNEIKVTWVPEKYKTLFD 290

RESULT 2

US-09-830-230A-362
; Sequence 362, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 362
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-362

Query Match 95.0%; Score 1451; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 6e-142;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 CDEKSSKNLKVIGYVNWGGTAATNVLKVFVKMGYNABIFSVTTSIMYQLASGKI 75
Db 1 CDEKSSKNLKVIGYVNWGGTAATNVLKVFVKMGYNABIFSVTTSIMYQLASGKI 60
QY 76 DGTGSSVWPTADKYFEKTKLTFVDLGANYEGTIQGFVVPSPVPISSISLKGKDGKPKN 135
Db 61 DGTGSSVWPTADKYFEKTKLTFVDLGANYEGTIQGFVVPSPVPISSISLKGKDGKPKN 120
QY 136 KMGIDAGAGTQVTEQALNYGLSKEYELVPSSSVMLASLSSSKRNEWILVPLMKPH 195
Db 121 KMGIDAGAGTQVTEQALNYGLSKEYELVPSSSVMLASLSSSKRNEWILVPLMKPH 180
QY 196 WAFSRYDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVYVDFHFWSDLLILPLMDKN 255
Db 181 WAFSRYDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVYVDFHFWSDLLILPLMDKN 240
QY 256 DKEPGKEYRNAVEFVEKNEIKVTWVPEKYKTLFD 290
Db 241 DKEPGKEYRNAVEFVEKNEIKVTWVPEKYKTLFD 275

RESULT 3

US-09-107-532A-5217
; Sequence 5217, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESSES: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-8277
; TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 577 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...577
; SEQUENCE DESCRIPTION: SEQ ID NO: 5217:
US-09-107-532A-5217

Query Match 32.3%; Score 493; DB 2; Length 577;
Best Local Similarity 37.0%; Pred. No. 3.7e-42;
Matches 97; Conservative 65; Mismatches 92; Indels 8; Gaps 4;
QY 21 SSKNLKSVKIGYVNWGGTAATNVLKVFVKMGYNABIFSVTTSIMYQLASGKIDGTYS 80
Db 320 SAKETKRLNSYVEMDEVASTNVVGEVLKQMGYDVTMTPLDINSIMKSVSNGESDAMVS 379
QY 81 SWVPADKFYFEKTKLTFVDLGANYEGTIQGFVVPSPVPISSISLKGKDGKFKNMIGI 140
Db 380 AWLPKTHGSOYAQYKQVEDLGANLTGAKVGLVAPYMDVNSIDELTQAGK--KIIGI 436
QY 141 DAGAGTQVTEQALNYGLSKEYELVPSSSVMLASLSSSKRNEWILVPLMKPHAFSR 200
Db 437 EPGAGVWTAENTIQKYNLNDKWKVETSSSGAMTVLALQAIKKHEPIVVTGTWPHWPAK 496
QY 201 YDIKFLDDPDLIMGGIESVHTLVRLGLENDDFDAYVYVDFHFWSD--DLILPLMD-KNDKE 258
Db 497 YDLKYLEDPENGMGSEEQIHTWVRKGLKEDQPEAYKYKVLNDHFHSEKMEKVMLEINNGKD 556
QY 259 PGKEYRNAVEFVEKNEIKVTW 280
Db 557 P---QQAADWKIKENQELVESW 575

RESULT 4

US-09-902-540-15605
; Sequence 15605, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 4.64138 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-8
Perfect score: 1527
Sequence: 1 MYKLFPIIFLSCDEKK.....EKNKEIVKTWPEKYKTLFD 290

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08 NEW PUB.pap.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06 NEW PUB.pap.*
- 3: /cgn2_6/ptodata/1/pubpaa/US07 NEW PUB.pap.*
- 4: /cgn2_6/ptodata/1/pubpaa/US09 NEW PUB.pap.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10 NEW PUB.pap.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11 NEW PUB.pap.*
- 7: /cgn2_6/ptodata/1/pubpaa/US12 NEW PUB.pap.*
- 8: /cgn2_6/ptodata/1/pubpaa/US60 NEW PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262	17.2	296	US-10-510-386-58	Sequence 58, Appl
2	98.5	6.5	663	US-11-196-475-78	Sequence 78, Appl
3	95	6.2	372	US-10-793-626-182	Sequence 182, Appl
4	92.5	6.1	700	US-11-196-475-74	Sequence 74, Appl
5	91.5	6.0	348	US-11-129-143-72	Sequence 72, Appl
6	91.5	6.0	663	US-11-196-475-70	Sequence 70, Appl
7	89	5.8	693	US-11-196-475-68	Sequence 68, Appl
8	86	5.6	568	US-10-793-626-2482	Sequence 2482, Ap
9	86	5.6	1145	US-10-793-626-1432	Sequence 1432, Ap
10	85.5	5.6	708	US-11-196-475-76	Sequence 76, Appl
11	85	5.6	585	US-11-074-176-190	Sequence 190, Appl
12	84.5	5.5	687	US-11-074-176-260	Sequence 260, Appl
13	82	5.4	429	US-10-793-626-3174	Sequence 3174, Ap
14	82	5.4	693	US-11-196-475-72	Sequence 72, Appl
15	81.5	5.3	700	US-11-196-475-66	Sequence 66, Appl
16	81.5	5.3	2516	US-10-647-956A-2	Sequence 2, Appli
17	81	5.3	271	US-11-052-554A-225	Sequence 225, App
18	81	5.3	567	US-10-485-517-216	Sequence 216, App
19	81	5.3	743	US-10-485-517-351	Sequence 351, App
20	81	5.3	817	US-10-793-626-2948	Sequence 2948, Ap
21	81	5.3	877	US-10-485-517-200	Sequence 200, App
22	81	5.3	2710	US-11-051-453-41	Sequence 41, Appl
23	80.5	5.3	255	US-10-485-517-360	Sequence 360, App
24	80.5	5.3	392	US-10-793-626-194	Sequence 194, App
25	79.5	5.2	175	US-10-793-626-1928	Sequence 1928, Ap

ALIGNMENTS

RESULT 1
US-10-510-386-58
; Sequence 58, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510.386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 58
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-58

Query Match 17.2%; Score 262; DB 6; Length 296;
Best Local Similarity 35.2%; Pred. No. 1.7e-15;
Matches 57; Conservative 35; Mismatches 64; Indels 6; Gaps 3;
QY 121 SSISELKGKDKFRKVMIGIDAGAGTQIVTQALNYGLSKEYELVPSESVMLASLDSS 180
Db 24 SNNNENASVGQVNYKITGIDPGAGIMNATDQALXDYDLK-WTVTSGSSSANTAAKKA 82
QY 161 IKNENWILVPLKHPWAFSRVDIKELDDPLIMGIESVHTLVRLGLENDPDFAYVDFH 240
Db 83 YDKDPPIITGTWPHMFAKYDLKYLKDPKSGYDAEIEHTVTRKGFDDHFGANKLLSQ 142
QY 241 FVMSDDL--LPLMDNDKEPGKEYRNAVEFVEKKNKEIVKTW 280
Db 143 FSWTDDMGVNLAVQEGKKPEEA---AADFVKHQDLVKKW 161

RESULT 2
US-11-196-475-78
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Comes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.

```
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631,1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-11-196-475-78

Query Match      6.5%; Score 98.5; DB 7; Length 663;
Best Local Similarity 19.9%; Pred. No. 0.53;
Matches 68; Conservative 58; Mismatches 97; Indels 119; Gaps 16;

QY 1 MYKFLFLFIIFMF---LSCDE--KKSKNKLKSVKGYVNWGGETAATNVLVKVFVKMGYN 55
Db 4 MLLIFSFVLFLNGFLNAREVDKELKDFVNMDFVNYKGPYDSTNTYEQI---VGIG 60
QY 56 AEIFSVTTSIMQYVLASGKIDGTSSVWPTADKFYVEKLTKEFVD-----LGA 103
Db 61 -----EFLARPLINSNS-----SYGKYFVNRFDIDQDKKASVDIFSIGS 102
QY 104 -----NYEGTIQGFVVPSYVPISSISELKSGDKFKFNKMGIDAGAGTQIVTEQALN 155
Db 103 KSELDLSILNRLITGLYLMKSPDYERSAELIAKAITIYNAVYRGD-----LD 150
QY 156 YGLSKEYELVPSSESVMLASLDSSIKRNEW-----ILVPLMKPHWAFSRDYIKFLDDPD 210
Db 151 YY---KEFYIASLKLTKENAGLSRVYSQWAGKTQIFIFLKK-----N 191
QY 211 LIMGGIES---VHTLV-----RLGLEND-----DFDAYVYV-FDHF 241
Db 192 ILSGNVEISDIDSLVTDKVAALLSENGSVNFARIDITIOGETHKADQDKIDIELDNF 251
QY 242 YWSDDLILPLMDK-----NDKEPGKEYRNAVEFVEKNKE 275
Db 252 HESDSNITETIENLRDLEKATDDEHKKIESQVDAKKQKE 293

RESULT 3
US-10-793-626-182
; Sequence 182, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 182
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
```

```
US-10-793-626-182

Query Match      6.2%; Score 95; DB 6; Length 372;
Best Local Similarity 19.7%; Pred. No. 0.49;
Matches 54; Conservative 37; Mismatches 75; Indels 108; Gaps 12;

QY 39 TAATNVLVKVFVKMGYNAEIFSFTTSMYQYVLASGKIDGTSSVW-PTADKFYVEKLTKE 97
Db 92 TFTPVMVSLTEQFNINIPATVMTDYRMH-----KNWITPYSQRYVATKDTK 140
QY 98 --FVDLGANYEGTIQGFVVPSYVPISSISELKSGDKFKFNK-----IGIDAGAGTQ 147
Db 141 DDFIEAG-----VPASYIKVTGIP-----IADKFEESIDKEWLSQQHLDPSPKPTI 186
QY 148 IVTEQALNYGLSKEYELV-----PSSSVMLASLDSSIKRNEWILVPLMKPHWAFS 199
Db 187 LMSAGA---FGVSKGFDYMINNILEKSPNSQVWVICGRSKELKRS-----228
QY 200 RYDIKFLDDPDLIM-----GGIESVHTLVRLGLENDDFDAYVYV 238
Db 229 -LKAKFKDNPSVILIGYTNHNMWASSQLMTKPGGITISEGLR-----273
QY 239 DHFYWSDDLILPLMDKNDKEPGKEYRNAVEFVEK 272
Db 274 -----CIPMIPLN-PAPQSELENAYYFESK 297

RESULT 4
US-11-196-475-74
; Sequence 74, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631,1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-11-196-475-74

Query Match      6.1%; Score 92.5; DB 7; Length 700;
Best Local Similarity 19.6%; Pred. No. 1.9;
Matches 67; Conservative 58; Mismatches 98; Indels 119; Gaps 16;

QY 1 MYKFLFLFIIFMF---LSCDE--KKSKNKLKSVKGYVNWGGETAATNVLVKVFVKMGYN 55
Db 4 MLLIFSFVLFLNGFLNAREVDKELKDFVNMDFVNYKGPYDSTNTYEQI---VGIG 60
QY 56 AEIFSVTTSIMQYVLASGKIDGTSSVWPTADKFYVEKLTKEFVD-----LGA 103
Db 61 -----EFLARPLINSNS-----SYGKYFVNRFDIDQDKKASVDIFSIGS 102
QY 104 -----NYEGTIQGFVVPSYVPISSISELKSGDKFKFNKMGIDAGAGTQIVTEQALN 155
```



```
Db 103 KSELDILNRLITCYLMKSFYERSSAELIAKAITIYNAVVRGD-----LD 150
QY 156 YTGSLKEYELVPSSEVMASLDSSIKRNEW-----ILVPLWKPWFASRYDIKFLDDPD 210
Db 151 YY---KEFYIASLSLTKENAGLSRVYSQWAGTKQIFPLKK-----N 191
QY 211 LIMGIES---VHTLV-----RLGLEND-----DPDAYVY-PDHP 241
Db 192 ILSGNVEDIDISLTVTKVVAALLSENGSVNFARDITDIQETHKADQDKIDIELONI 251
QY 242 YMSDDLILFLMDK-----NDKEPGKEYRNAVEFVEKNKE 275
Db 252 HESDSNITETIENLRDLEKATDEEHKKIEISQVDAKKQKE 293

RESULT 5
US-11-129-143-72
; Sequence 72, Application US/11129143
; Publication No. US20050266518A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMBELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne P.
; APPLICANT: YELISEEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/11/129,143
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72
; LENGTH: 348
; TYPE: PRM
; ORGANISM: Thermoplasma acidophilum
US-11-129-143-72

Query Match 6.0%; Score 91.5; DB 7; Length 348;
Best Local Similarity 24.6%; Pred. No. 0.89;
Matches 68; Conservative 29; Mismatches 100; Indels 79; Gaps 14;

QY 23 KULKSVKIGYVWGGETAATNV---LKVPFKMGYNAEIFSTTSIMYOYLASGKIDTV 79
Db 52 KKLKPFMISSMTGCAETAKNIRNLAVAAERFGIMGVSMRAIV-----DRSTIEDTY 106
QY 80 S-----SWVP-----TADK-----FYEKLKTKFVDLGANYEGTIOGFVV 114
Db 107 SVINSHVPLKTIANTGAPOLVRQDKDQVNSNRDIAYIDLIKADFLAVHFN---LQEMVQ 163
QY 115 P-----SYVPISISSELKKG-----GDKFKNMIG---IDAG-----AGTQIVTEQA 153
Db 164 PGSDRNSKGVDRIDKDLGSGFNIIAKETSGSGFSRRTAELIDAGVKAEVSGVGTTPAA 223
QY 154 LNYGLSKKEYEL-----VPSSESVMLAS-----LDSSIKRNEWIL-----VPLW 192
Db 224 VEYTRARKENNLKMRIGETFWNWGIPSPASVYVCSDLAPVIGSGRLGLDLAKAIAMG 283
QY 193 KHWAFSRVDIKFLD-DPDLIMGIESVHTLVRLGL 227
Db 284 ATAGGFARSLKADATDPEMLMKNIELIQREFRVAL 319

RESULT 6
US-11-196-475-70
; Sequence 70, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
```

```
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 663
; TYPE: PRM
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match 6.0%; Score 91.5; DB 7; Length 663;
Best Local Similarity 20.6%; Pred. No. 2.1;
Matches 68; Conservative 56; Mismatches 111; Indels 95; Gaps 15;

QY 1 MYKLFLFFIIFMF---LSCDE--KKSSKNLKSVKIGYVNWGGETAATNVLKVPFKMGYN 55
Db 4 MLLPSPFLVFLNGPPLNAREVDKELKDFVNMDLFEVYKGPYDSTDTYEQIVGIGBFL 63
QY 56 AEIPSVTTSIMYOYLASGKIDTVSSVPTADKFFYEKLKTKFVDLGA-----NYEG 107
Db 64 ARPLNNSNSSY--GKY--FVNRFDQDK---KASVDIFSIGSKSELDLSILNLR 114
QY 108 TIQGFVPSYVPISISSELKKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKKEYELVP 167
Db 115 ILTYLMKSFYERSSAELIAKAITIYNAVVRGD-----LDYY---KEFYIEA 159
QY 168 SSESVMASLDSSIKRNEW-----ILVPLWKPWFASRYDIKFLDDPDILIMGIES---V 219
Db 160 SLKSLTKENAGLSRVYSQWAGTKQIFPLKK-----NILSGNVEDIDI 203
QY 220 HTLV-----RLGLEND-----DPDAYVY-FDHFVWSDLLPLMD 253
Db 204 DSLVTDKVVAAALLSENGSVNFARDITDIQETHKADQDKIDIELDNFHESDSNITETIE 263
QY 254 K-----NDKEPGKEYRNAVEFVEKNKE 275
Db 264 NLRDLEKATDEEHKKIEISQVDAKKQKE 293

RESULT 7
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-68

Query Match      5.8%; Score 89; DB 7; Length 693;
Best Local Similarity 20.0%; Pred. No. 3.7;
Matches 63; Conservative 54; Mismatches 136; Indels 62; Gaps 12;

QY 1 MYKLFLLFFLFLSCD-----EKSSKNLKSVKIGYVNWGGETAATNVLVKVPFEK 51
Db 1 MKKLLIFLFFFL-LSLNGPLNSREVDKEKJDFVNMDFVNYKGPYDSTNTYSQIVGI 59

QY 52 MGYNAE--IFSVTTSIMYQLASGKIDGTVSSWVPTADKFYFEKLTCKFVDLGA----- 103
Db 60 GFLARPLNSNSNIY-----GKY--FINRFIDDDK-----KASVDVFSIGSRSQLDS 108

QY 104 --NYEGTIOGFVPSVPSISSELSKGGDKFKNMKMGIDAGAGTQIVTEQALN 157
Db 109 ILNRLRLTYLKSIDYERSAELIAKVTITHNAVYRGLNVYKEVYTEAALKSLTKEN 168

QY 158 -GLSKEYE-----LVPSSSEVMSLASDSSIKRNEWILVPLKPHWAFSYDIKFLDD 208
Db 169 AGLSRVYSQWAGKTQIFILKKNILSGKVESDIDSLVTDKVAALJSENEAGVNFARD 228

QY 209 POLIMGGIESVHTLVRLGLENDDFDAYVDFHYWSDDLLILPLMDK-----NDKEPG 260
Db 229 ITDIQETHKA-----DQDKID--IELDNVHKSDSNITETIENLRDQLEKATDEHR 278

QY 261 KEYNNAVEFVEKNKE 275
Db 279 KEIESQVDAKKQKE 293

RESULT 8
US-10-793-626-2482
; Sequence 2482, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2482
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2482

Query Match      5.6%; Score 86; DB 6; Length 568;
Best Local Similarity 19.6%; Pred. No. 5.1;
Matches 70; Conservative 59; Mismatches 124; Indels 104; Gaps 15;

QY 8 FIIFMLSCDEKKSKNLKSVKIGYVNWGGETAATNVLVKVPFKMGYNAEIFSVT----- 62
Db 246 FLVFM---DQHRESTDYVNRITQIFITDAQYIYIQNMQSILDTLQINEEQFEAYIQPQ 302

QY 63 -----TSIMYQYLASGKIDGTVSSWVPTAD-----KFYFEKLTCKFVDLGA 103
Db 303 QNQEVSAAQLLNSDNQFLNFKQKRDILDNAYIMTYDMRESLRNYLESMATDF----- 357

QY 104 NYEGTIOGFV---VPSYVPISSISE-----LKGKDGKFKNMKMGIDAGAGTQIVTEQALN 155
Db 358 ----KVNGFFNKRKKKEEQIKRLNEATTQLQEKVQVQVQVQVQVQVQVQVQVQVQVQV 413

QY 156 YGLSKVELVPSSSEVMSLASDSSIKRNEWILVPLKPHWAFSYDIKFL-----DDP 209
Db 414 EKILNQEYDVVPSLIS-ELYQTQTSIS-NTYVL-----TFSEVIALKNNKKTENEST 463

QY 210 DLIMGGIESVHTLVRLGLEND--FDAY-----YVDFHYWSDDLLILPLMDK 254
Db 464 PLFEAVNHVQVNELSSDENEDRYEYRIELNTLKDSLTSNHYKHYY-----IHLDDS 517

QY 255 NDKEPG-----EYRNAVEFVEKNKEIVKTVWPEKYKTLFD 290
Db 518 LDKLIGRTETHPELKQENSTAYHRKHETQHRN--EFVTSNQDKRALDIVKDVPLFD 572

RESULT 10
US-11-196-475-76
; Sequence 76, Application US/11196475

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 50.9692 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-8
Perfect score: 1527
Sequence: 1 MYKLFPLFIIPFMSLSCDEKK.....EKNKEIVKTVPEKYKTLFD 290

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1527	100.0	290	US-10-688-058-8	Sequence 8, Appli
2	1527	100.0	290	US-10-994-726-361	Sequence 361, App
3	1451	95.0	275	US-10-994-726-362	Sequence 362, App
4	468	30.6	91	US-10-688-058-6	Sequence 6, Appli
5	459	30.1	569	US-09-815-242-10503	Sequence 10503, A
6	442.5	29.0	575	US-10-474-792-100	Sequence 100, App
7	433.5	28.4	320	US-10-501-282-1792	Sequence 1792, App
8	416.5	27.3	334	US-10-501-282-1796	Sequence 1796, App
9	416.5	27.3	352	US-10-501-282-1798	Sequence 1798, App
10	408	26.7	302	US-10-501-282-5510	Sequence 5510, App
11	406	26.6	303	US-10-501-282-1918	Sequence 1918, App
12	406	26.6	326	US-10-501-282-1920	Sequence 1920, App
13	370.5	24.3	871	US-10-156-761-14244	Sequence 14244, A
14	338.5	22.2	273	US-10-501-282-1794	Sequence 1794, App
15	235	15.4	312	US-09-815-242-5213	Sequence 5213, App
16	235	15.4	312	US-10-282-122A-43548	Sequence 43548, A
17	224	14.7	380	US-10-282-122A-49743	Sequence 49743, A
18	189	12.4	314	US-10-282-122A-57487	Sequence 57487, A
19	118	7.7	158	US-10-724-972A-5439	Sequence 5439, App
20	113	7.4	165	US-10-724-972A-5331	Sequence 5331, App
21	112.5	7.4	324	US-10-156-761-10967	Sequence 10967, A
22	108.5	7.1	266	US-09-765-272-14	Sequence 14, Appl
23	108.5	7.1	266	US-11-106-649-14	Sequence 14, Appl
24	106	6.9	333	US-10-282-122A-5335	Sequence 5335, A
25	101	6.6	1842	US-10-243-552-847	Sequence 847, App
26	101	6.6	1842	US-10-243-552-887	Sequence 887, App
27	100.5	6.6	570	US-10-369-493-5353	Sequence 5353, App

28	100.5	6.6	663	4	US-10-289-795-5	Sequence 5, Appli
29	100.5	6.6	663	5	US-10-762-665-5	Sequence 5, Appli
30	100	6.5	271	4	US-10-282-122A-76195	Sequence 76195, A
31	99	6.5	1184	4	US-10-282-122A-53497	Sequence 53497, A
32	98.5	6.5	663	4	US-10-369-100-78	Sequence 78, Appl
33	97.5	6.4	270	4	US-10-282-122A-52838	Sequence 52838, A
34	97.5	6.4	274	3	US-09-971-536-58	Sequence 58, Appli
35	97	6.4	265	5	US-10-514-054-2	Sequence 2, Appli
36	97	6.4	348	4	US-10-282-122A-47221	Sequence 47221, A
37	97	6.4	437	4	US-10-282-122A-61263	Sequence 61263, A
38	96	6.3	565	5	US-10-732-923-1913	Sequence 1913, Ap
39	96	6.3	656	4	US-10-424-599-173577	Sequence 173577, A
40	96	6.3	843	4	US-10-282-122A-53395	Sequence 53395, A
41	95	6.2	221	5	US-10-617-320-2796	Sequence 2796, Ap
42	95	6.2	391	4	US-10-724-972A-5743	Sequence 5743, Ap
43	94.5	6.2	536	4	US-10-369-493-6866	Sequence 6866, Ap
44	94.5	6.2	727	4	US-10-437-963-103577	Sequence 103577, A
45	94	6.2	271	4	US-10-282-122A-55826	Sequence 55826, A

ALIGNMENTS

RESULT 1
US-10-688-058-8
; Sequence 8, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:00305
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 8
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-8

Query Match	100.0%;	Score 1527;	DB 5;	Length 290;
Best Local Similarity	100.0%;	Pred. No. 2.2e-134;		
Matches 290;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MYKLFPLFIIPFMSLSCDEKKSSKNLKVIGYVNWGGTAATNVLKVVFERKMGYNAEIFS	60	
Db	1	MYKLFPLFIIPFMSLSCDEKKSSKNLKVIGYVNWGGTAATNVLKVVFERKMGYNAEIFS	60	
QY	61	VTTSIMQYLAGKIDGTSSVWPTADKFFYEKLTKFVDLCANYEGTIQGFVPSYVPI	120	
Db	61	VTTSIMQYLAGKIDGTSSVWPTADKFFYEKLTKFVDLCANYEGTIQGFVPSYVPI	120	
QY	121	SSISELKGKDKFKNMIGIDAGAGTQIVTQALNYYGLSKYEYLVPSSESVMLASLDSS	180	
Db	121	SSISELKGKDKFKNMIGIDAGAGTQIVTQALNYYGLSKYEYLVPSSESVMLASLDSS	180	
QY	181	IKRNEWILVPLWKPHWAFSPRYDIKFLDDPDLIMGGIESVHTLVRLGLENDPDFDAYVFDH	240	
Db	181	IKRNEWILVPLWKPHWAFSPRYDIKFLDDPDLIMGGIESVHTLVRLGLENDPDFDAYVFDH	240	
QY	241	FYWSDDLLPLMDKNDKEPGKEYNVAFVEKNKEIVKTVPEKYKTLFD	290	
Db	241	FYWSDDLLPLMDKNDKEPGKEYNVAFVEKNKEIVKTVPEKYKTLFD	290	

RESULT 2
US-10-994-726-361
; Sequence 361, Application US/10994726

Publication No. US20050147999A1
 GENERAL INFORMATION:
 APPLICANT: Ni et al.
 TITLE OF INVENTION: Lyme Disease Vaccines
 FILE REFERENCE: PB481D1
 CURRENT APPLICATION NUMBER: US/10/994,726
 CURRENT FILING DATE: 2004-11-23
 PRIOR APPLICATION NUMBER: 09/830,230
 PRIOR FILING DATE: 2001-04-24
 PRIOR APPLICATION NUMBER: PCT/US98/12718
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/057,483
 PRIOR FILING DATE: 1997-09-03
 PRIOR APPLICATION NUMBER: 60/053,344
 PRIOR FILING DATE: 1997-07-22
 PRIOR APPLICATION NUMBER: 60/053,377
 PRIOR FILING DATE: 1997-07-22
 PRIOR APPLICATION NUMBER: 60/050,359
 PRIOR FILING DATE: 1997-06-20
 NUMBER OF SEQ ID NOS: 756
 SOFTWARE: PatentIn Ver. 3.3
 SEQ ID NO 361
 LENGTH: 290
 TYPE: PRT
 ORGANISM: Borrelia burgdorferi
 US-10-994-726-361

Query Match 100.0%; Score 1527; DB 5; Length 290;
 Best Local Similarity 100.0%; Pred. No. 2.2e-134;
 Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYKFLFFIIFWFLSCDEKSSKNLKVIGYVNWGGETAATNVLKVPKMGYNALIFS 60
 DB 1 MYKFLFFIIFWFLSCDEKSSKNLKVIGYVNWGGETAATNVLKVPKMGYNALIFS 60

QY 61 VTTSIMYQYLAGSKIDGTSSWVPTADKFYEEKLTKTFVDLGANYEGTTIGGFVPSYVPI 120
 DB 61 VTTSIMYQYLAGSKIDGTSSWVPTADKFYEEKLTKTFVDLGANYEGTTIGGFVPSYVPI 120

QY 121 SSISELKGKDKPKKMGIGIDAGAGTQIVTEQALNYGLSKYELVPSSSVMLASLSS 180
 DB 121 SSISELKGKDKPKKMGIGIDAGAGTQIVTEQALNYGLSKYELVPSSSVMLASLSS 180

QY 181 IKRNEWILPLKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPFDAYYVFDH 240
 DB 181 IKRNEWILPLKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPFDAYYVFDH 240

QY 241 FYWSDDLILPLMDKNDKEPKGYRNAVEFVEKKNKEIVKTWVPEKYKTLFD 290
 DB 241 FYWSDDLILPLMDKNDKEPKGYRNAVEFVEKKNKEIVKTWVPEKYKTLFD 290

RESULT 3
 US-10-994-726-362
 Sequence 362, Application US/10994726
 Publication No. US20050147999A1
 GENERAL INFORMATION:
 APPLICANT: Ni et al.
 TITLE OF INVENTION: Lyme Disease Vaccines
 FILE REFERENCE: PB481D1
 CURRENT APPLICATION NUMBER: US/10/994,726
 CURRENT FILING DATE: 2004-11-23
 PRIOR APPLICATION NUMBER: 09/830,230
 PRIOR FILING DATE: 2001-04-24
 PRIOR APPLICATION NUMBER: PCT/US98/12718
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/057,483
 PRIOR FILING DATE: 1997-09-03
 PRIOR APPLICATION NUMBER: 60/053,344
 PRIOR FILING DATE: 1997-07-22
 PRIOR APPLICATION NUMBER: 60/053,377
 PRIOR FILING DATE: 1997-07-22
 PRIOR APPLICATION NUMBER: 60/050,359

PRIOR FILING DATE: 1997-06-20
 NUMBER OF SEQ ID NOS: 756
 SOFTWARE: PatentIn Ver. 3.3
 SEQ ID NO 362
 LENGTH: 275
 TYPE: PRT
 ORGANISM: Borrelia burgdorferi
 US-10-994-726-362

Query Match 95.0%; Score 1451; DB 5; Length 275;
 Best Local Similarity 100.0%; Pred. No. 2.6e-127;
 Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 CDEKSSKNLKVIGYVNWGGETAATNVLKVPKMGYNALIFSITTSIMYQYLAGSKI 75
 DB 1 CDEKSSKNLKVIGYVNWGGETAATNVLKVPKMGYNALIFSITTSIMYQYLAGSKI 60

QY 76 DGTSSWVPTADKFYEEKLTKTFVDLGANYEGTTIGGFVPSYVPISSISELKGKDKPKN 135
 DB 61 DGTSSWVPTADKFYEEKLTKTFVDLGANYEGTTIGGFVPSYVPISSISELKGKDKPKN 120

QY 136 KMIGIDAGAGTQIVTEQALNYGLSKYELVPSSSVMLASLSSIKRNEWILVPLWPKH 195
 DB 121 KMIGIDAGAGTQIVTEQALNYGLSKYELVPSSSVMLASLSSIKRNEWILVPLWPKH 180

QY 196 WAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPFDAYYVFDHFYWSDDLILPLMDKN 255
 DB 181 WAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPFDAYYVFDHFYWSDDLILPLMDKN 240

QY 256 DKEPKGYRNAVEFVEKKNKEIVKTWVPEKYKTLFD 290
 DB 241 DKEPKGYRNAVEFVEKKNKEIVKTWVPEKYKTLFD 275

RESULT 4
 US-10-688-058-6
 Sequence 6, Application US/10688058
 Publication No. US20050058661A1
 GENERAL INFORMATION:
 APPLICANT: SYKES, KATHRYN F.
 APPLICANT: HALE, KATHERINE S.
 APPLICANT: JOHNSTON, STEPHEN A.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
 TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
 TITLE OF INVENTION: BORRELIA
 FILE REFERENCE: MCRO:003US
 CURRENT APPLICATION NUMBER: US/10/688,058
 CURRENT FILING DATE: 2003-10-17
 NUMBER OF SEQ ID NOS: 141
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 6
 LENGTH: 91
 TYPE: PRT
 ORGANISM: Borrelia burgdorferi
 US-10-688-058-6

Query Match 30.6%; Score 468; DB 5; Length 91;
 Best Local Similarity 98.9%; Pred. No. 6.9e-36;
 Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 109 IQGFVVPSYVPISSISELKGKDKPKKMGIGIDAGAGTQIVTEQALNYGLSKYELVPS 168
 DB 1 IQGFVVPSYVPISSISELKGKDKPKKMGIGIDAGAGTQIVTEQALNYGLSKYELVPS 60

QY 169 SESVMLASLSSIKRNEWILVPLWPKHWAFS 199
 DB 61 SESVMLASLSSIKRNEWILVPLWPKHWAFS 91

RESULT 5
 US-09-815-242-10503
 Sequence 10503, Application US/09815242
 Patent No. US20020061569A1

GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cart, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10503
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10503

Query Match 30.1%; Score 459; DB 3; Length 569;
Best Local Similarity 35.5%; Pred. No. 6e-34;
Matches 93; Conservative 59; Mismatches 102; Indels 8; Gaps 4;
QY 21 SKNLSKVKIGYNNWGGETAATNLVKVPEKMGYNABIFSVTTSIMYQVYLASGKIDGTVS 80
DB 312 TTNDKQISLSYVEMDEVSTAVTHVAEVLKMGYDVKVKTPLDNLAMWESVAKGETDAMVG 371
QY 81 SWPTADKPYEKLTKFVDLGNAYEGTIGQFVVPSPVPISSISELKGDKFKNMIGI 140
DB 372 AMLPGTHAEQYKQYKLDLGNLKGAKLGIWPSYMDVDSIEDLSQAGK---KITGI 428
QY 141 DAGAGTQIVTEQALNYGLSKEYELVPSSSEYMLASLSDSSIKRNEWILVPLKPHWAFSR 200
DB 429 EFGAGVVAAEKTKAYPNLKDWSVETSSSGAMTVALGQAIKKNEDIVITGWSPHWFAK 488
QY 201 YDIKFLDDPLIMGGIESVHTLVRLGLENDPDFAYVDFHFW--SDDLPLPLMDKNDKEP 259
DB 489 YDLKYLADPKGTMGGEAHTHARQGLKEDQPEAYKVLNDFHFWTTKDMESVMLEINE--- 545
QY 260 GKEYRNAV--EFVEKKNKEIVKTW 280
DB 546 GKDPQEAARDWVDSHKDQVAEW 567

RESULT 6
US-10-474-792-100
; Sequence 100, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792

; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 100
; LENGTH: 575
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-100
Query Match 29.0%; Score 442.5; DB 5; Length 575;
Best Local Similarity 35.8%; Pred. No. 2.1e-32;
Matches 92; Conservative 56; Mismatches 102; Indels 7; Gaps 5;
QY 26 KSVKIGYNNWGGETAATNLVKVPEKMGYNABIFSVTTSIMYQVYLASGKIDGTVSWSVPT 85
DB 322 ETNIAIYQWDSEVASTHVAEVLKNEGYHVTLTPLDNVAMVMTVANGNADPSTSAWLPV 381
QY 86 ADKPYEKLTKFVDLGNAYEGTIGQFVVPSPVPISSISELKGDKFKNMIGIDAGA 144
DB 382 THGQYQYKYSKLDLGNLKGAKLGIWPSYMDVDSIEDLSQADQ---KITGIEPCA 438
QY 145 GTQIVTEQALNYGLSKEYELVPSSSEYMLASLSDSSIKRNEWILVPLKPHWAFSRYDIK 204
DB 439 GIMAAQKTLKEYHNLSSELVAASTGAMTTSLDQAIKKKOPIVVTAWSPHWFAKYDIK 498
QY 205 FLDDPLIMGGIESVHTLVRLGLENDPDFAYVDFHFWSD--DLILPLMDKNDKEPGKEY 263
DB 499 YLKPKEIFGSTEININTIARKLKELPNVYKIIDKFHTWQKDMEAVALDIN-KGMSPE- 556
QY 264 RNAVEFVEKKNKEIVKTW 280
DB 557 AAKKWEANKSVSSW 573

RESULT 7
US-10-501-282-1792
; Sequence 1792, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AML00780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1792
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Alloiooccus otitidis
US-10-501-282-1792

Query Match 28.4%; Score 433.5; DB 5; Length 320;
Best Local Similarity 33.7%; Pred. No. 6.7e-32;
Matches 93; Conservative 62; Mismatches 110; Indels 11; Gaps 4;
QY 12 MFL-SCDEKSKSNLKS-----KIGVYNNWGGETAATNLVKVPEKMGYNABIFSVTTS 64
DB 16 LFLTACQGEDGAADVSVVDGEEIBLAYVDMWDSIATSHVIGEVLESIGLDVTLTSLDNA 75
QY 65 IMYQVYLASGKIDGTVSWSVPTADKPYEKLTKFVDLGNAYEGTIGQFVVPSPVPISSIS 124
DB 76 VMWQSVANGADAMVSAWLPHTHGQEAEBYGDQMDHVGNTLEGAKIGLGVPTIMEVDSTA 135

RESUME 13
US-106-761-14244
Sequence 14244, Application US/1056761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE

Db 205 PNNLQIDMTYLTGSEVDVFGPDEGAATVSTMTAAGYQAQCNCNVARLLHNLRFSSAQV----- 260

Qy 242 YWSDDLILPLMDKNDKEPGKGYRNAVEFVEKNKEIVKTVV 281

Db 261 ---SQVMAPIILDRT-----QPLDAAQWLKANPEPLKAWL 292

Search completed: January 24, 2006, 20:54:30
Job time : 51.9692 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 62.3148 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-8

Perfect score: 1527

Sequence: 1 MYKFLFLPIIFMFLSCDEKK.....EKNKEIKVTWVPEKVKILFD 290

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1527	100.0	290	051169 BORBU	051169 borrelia bu
2	1454.5	95.3	294	Q662L9 BORBU	Q662L9 borrelia bu
3	510	33.4	285	Q729Q3 DESVH	Q729Q3 desulfovibr
4	505	33.1	315	Q8TNY0 METAC	Q8TNY0 methanosarc
5	494.5	32.4	289	Q926L5 LISIN	Q926L5 listeria in
6	478.5	31.3	277	Q8A6X6 BACTIN	Q8A6X6 bacteroides
7	463.5	30.4	306	Q8U4S3 METMA	Q8U4S3 methanosarc
8	462	30.3	282	Q5WJD4 BACSK	Q5WJD4 bacillus cl
9	460.5	30.2	298	Q891M9 CLOTE	Q891M9 clostridium
10	459	30.1	569	Q830X6 ENTFA	Q830X6 enterococcu
11	451	29.5	573	Q9KIF6 9LACT	Q9KIF6 lactococcus
12	451	29.5	573	Q7DAU8 LACLA	Q7DAU8 lactococcus
13	447	29.3	297	Q8EXX8 OCEIH	Q8EXX8 oceanobacil
14	442.5	29.0	561	Q8K8R4 STRP3	Q8K8R4 streptococc
15	442.5	29.0	575	Q879N8 STRP3	Q879N8 streptococc
16	442.5	29.0	575	Q9A1M9 STRPY	Q9A1M9 streptococc
17	440.5	28.8	575	Q5XE24 STRP6	Q5XE24 streptococc
18	439	28.7	575	Q8XQ4 STRA5	Q8XQ4 streptococc
19	439	28.7	575	Q8E3C3 STRA3	Q8E3C3 streptococc
20	438	28.7	285	Q9HZ04 PSEAE	Q9HZ04 pseudomonas
21	436.5	28.6	575	Q8P2S5 STRP8	Q8P2S5 streptococc
22	433.5	28.4	287	Q4LV54 BURK	Q4LV54 burkholderi
23	428.5	28.1	573	Q9RQ05 9LACT	Q9RQ05 lactococcus
24	427.5	28.0	283	Q4KFZ7 PSEF5	Q4KFZ7 pseudomonas
25	424.5	27.8	576	Q8DU85 STRMU	Q8DU85 streptococc
26	418	27.4	287	Q63MU2 BURPS	Q63MU2 burkholderi
27	418	27.4	287	Q62DH6 BURMA	Q62DH6 burkholderi
28	416.5	27.3	285	Q63AG1 BACCZ	Q63AG1 bacillus ce
29	416	27.2	284	Q64R03 BACFR	Q64R03 bacteroides
30	415.5	27.2	295	Q6LYW9 METMP	Q6LYW9 methanococc
31	415	27.2	284	Q5LAK8 BACFN	Q5LAK8 bacteroides

RESULT 1
051169 BORBU PRELIMINARY; PRT; 290 AA.
AC 051169;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE Glycine betaine, L-proline ABC transporter, glycine/betaine/L-proline-binding protein (PROX).
GN OrderedLocNames=BB0144;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A., Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K., RA Winn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D., RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J., RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D., RA Gocayne J.D., Weidman J.F., Uterback T.R., Watthey L., McDonald L.A., RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K., RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi";
RL Nature 390:580-586 (1997).
DR EMBL; AB001125; AAC66525.1; -; Genomic_DNA.
DR PIR; H70117; H70117.
DR TIGR; BB0144; -.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 290 AA; 33263 MW; 43779CF0911FBBDS CRC64;

Query Match 100.0%; Score 1527; DB 2; Length 290;
Best Local Similarity 100.0%; Pred. No. 3.2e-108;
Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYKFLFLPIIFMFLSCDEKKSKNLKSVKIGYVNWGGTAAATNLVKVFFKMGYNNAETFS 60
DB 1 MYKFLFLPIIFMFLSCDEKKSKNLKSVKIGYVNWGGTAAATNLVKVFFKMGYNNAETFS 60
QY 61 VTTSIMYQYLASGKIDGTGVSSVWPTADKFFYYEKLKTKFVLDGANYEGTIQGFVWVSYPVI 120
DB 61 VTTSIMYQYLASGKIDGTGVSSVWPTADKFFYYEKLKTKFVLDGANYEGTIQGFVWVSYPVI 120
QY 121 SSISELKGKDGKFNKMGIGIDAGAGTQIVTQALNYGLSKYEYLVPSSESVMASLDSS 180

```
Db 121 SSISELKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKEYELVPSSSEVMLASLDSS 180
Qy 181 IKRNEWILVPLMKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPDAYVVDH 240
Db 181 IKRNEWILVPLMKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPDAYVVDH 240
Qy 241 FWYSDDLILPLMDKNDKPGKEYNRAVEFVEKNKEIVKTVWPEKYKTLFD 290
Db 241 FWYSDDLILPLMDKNDKPGKEYNRAVEFVEKNKEIVKTVWPEKYKTLFD 290

RESULT 2
Q662L9 BORGA
ID Q662L9 BORGA PRELIMINARY; PRT; 294 AA.
AC Q662L9 BORGA PRELIMINARY; PRT; 294 AA.
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DE Glycine betaine, L-proline ABC transporter, binding protein.
GN Name=proX; OrderedLocusNames=BG0144;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Stuehn J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000013; RAU07002.1; -; Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 294 AA; 33680 MW; AFA81026194ABE12 CRC64;

Query Match 95.3%; Score 1454.5; DB 2; Length 294;
Best Local Similarity 93.5%; Pred. No. 1.1e-102;
Matches 272; Conservative 13; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MYKFLFPIIFMF-LSCDEKSKNLSKVTKGVNMGGETAATNVLKVVFEXMGYNAEIF 59
Db 4 IYKFLFGFIIFLFLSCDEKSKNLSKVTKGVNMGGETAATNVLKVVFEXMGYNAEIF 63
Qy 60 SVTTSIMYQYLASGKIDGTSSWVPTADKPYEKLTKFVDLGANYEGTIQGFVVPYVP 119
Db 64 SVTTSIMYQYLATGKVDGTSSWVPTADKPYEKLTKFVDLGANYEGTIQGFVVPYVP 123
Qy 120 ISSISELKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKEYELVPSSSEVMLASLDS 179
Db 124 ISSISELKGDKFKNMIGIDAGAGTQIVTEQALDYYGLSKEYELVPSSSEVMLASLDS 183
Qy 180 SKRNEWILVPLMKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPDAYVVDH 239
Db 184 AIKRNWILVPLMKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPDAYVVDH 243
Qy 240 HFYWSDDLILPLMDKNDKPGKEYNRAVEFVEKNKEIVKTVWPEKYKTLFD 290
Db 244 HFYWSDDLILPLMDKNDKPGKEYNRAVEFVEKNKEIVKTVWPEKYKTLFD 294

RESULT 3
Q729Q3 DESVH
ID Q729Q3 DESVH PRELIMINARY; PRT; 285 AA.
AC Q729Q3 DESVH PRELIMINARY; PRT; 285 AA.
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Glycine/betaine/L-proline ABC transporter, periplasmic-binding
```

```
protein.
DE OrderedLocusNames=DVU2297;
GN Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15077118; DOI=10.1038/nbt959;
RA Heidelberg J.F., Seshadri R., Haveman S.A., Hemme C.L., Paulsen I.T.,
RA Kolonay J.F., Eisen J.A., Ward N.L., Methe B.A., Brinkac L.M.,
RA Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S., Madupu R.,
RA Nelson W.C., Sullivan S.A., Fouts D.E., Haft D.H., Selengut J.,
RA Peterson J.D., Davidsen T.M., Zafar N., Zhou L., Radune D.,
RA Dimitrov G., Hance M., Tran K., Khouri H.M., Gill J., Uitterback T.R.,
RA Feldblyum T.V., Wall J.D., Voordouw G., Fraser C.M.;
RT "The genome sequence of the anaerobic, sulfate-reducing bacterium
RT Desulfovibrio vulgaris Hildenborough.";
RL Nat. Biotechnol. 22:554-559(2004).
DR EMBL; AE017316; AAS96770.1; -; Genomic_DNA.
DR TIGR; DVU2297; -.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 285 AA; 31240 MW; 7C18644EDDE6EE47 CRC64;

Query Match 33.4%; Score 510; DB 2; Length 285;
Best Local Similarity 37.0%; Pred. No. 1.1e-30;
Matches 105; Conservative 63; Mismatches 112; Indels 4; Gaps 4;

Qy 1 MYKFLFPIIFMF-LSCDEKSKNLSKVTKGVNMGGETAATNVLKVVFEXMGYNAEIF 58
Db 1 MKKIILLAVFLAFTGAQAASAKDAKPVRIAYVEWDCARATSNLVKAAIEDRLHRKVEL 60
Qy 59 FSVTTSIMYQYLASGKIDGTSSWVPTADKPYEKLTKFVDLGANYEGTIQGFVVPYVP 118
Db 61 LPVSAGAMMAVASGDVDTATVTLVTHGDLKRLGKVRDLGFLVDGLGAVPDXV 120
Qy 119 PISSISELKGDKFKNMIGIDAGAGTQIVTEQALNYYGLSKEYELVPSSSEVMLASLD 178
Db 121 TVNSIAELDAADAFGGRIIGIDPGAGLMRSEDAIKAYGL-KKQLVGGSGATWTAALA 179
Qy 179 SSIRKNEWILVPLMKPHWAFSRDYDKFLDDPDLIMGGIESVHTLVRLGLENDPDAYVVP 238
Db 180 DAIRKKEIVTVTASPHMFGRWQLKYLDDPKGALGGMESIHTVVRKGLDKDMPVFAFL 239
Qy 239 DHFYWSDDLIL-PLMDKNDKPGKEYNRAVEFVEKNKEIVKTVW 281
Db 240 DRFAYADTAQTLMAWNNEEGADPLTNARFMKEHPALVDWSL 283

RESULT 4
Q8TNYO METAC
ID Q8TNYO METAC PRELIMINARY; PRT; 315 AA.
AC Q8TNYO;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Glycine betaine/L-proline ABC transporter, solute-binding
DE protein.
GN Name=proX; OrderedLocusNames=MA2147;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
```


RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., DeArelano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Titrill A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Darrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.,
 RT "The genome of Methanococcus acetivorans reveals extensive metabolic
 and physiological diversity,"
 RL Genome Res. 12:532-542(2002).
 DR EMBL; AE010901; AAM05545.1; -; Genomic_DNA.
 DR GO; GO:0005488; F:binding; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007210; OpuAC_ABC.
 DR Pfam; PF04069; OpuAC; 1.
 KW Complete proteome.
 SQ SEQUENCE 315 AA; 34613 MW; 262494BFCT9649BA CRC64;
 Query Match 33.1%; Score 505; DB 2; Length 315;
 Best Local Similarity 41.6%; Pred. No. 3e-30;
 Matches 112; Conservative 42; Mismatches 107; Indels 8; Gaps 5;
 QY 19 KSSNNKSKVIGYVWGGTAATNVLKVPKMGY-NAEIFSVTTSIMYOVLASGKIDG 77
 DB 49 EAESELSPTVIGYVWGGTAATNVLKVPKMGY-NAEIFSVTTSIMYOVLASGKIDG 108
 QY 78 TVSSVPTADKFFYKELTKFKFDLGANYEGTIQGFVPSVYPISSISLKGKDGKFNKM 137
 DB 109 TTSAWLPYTHQSYWEAYGQDLSVQTNLEDCKGLVPSVYTIIDSIELNSEKDFNGQI 168
 QY 138 IGIDAGAGTQVTEQALNYGLSKYELVPSSESVMLASLSSIKRNEWILVPLKPHWA 197
 DB 169 IGIDPGAGIMQASSETAITDYDL--DMELVSGSSAAMTTSLKSIDSEEWTVVTLSPHWA 226
 QY 198 FSRVDIKFDLPDPLMGIESVHTLVRLGLENDDFDAYVDFHFWYS-DDLILPLMD-KN 255
 DB 227 FNRWDLKYLDPKGAAYGADHVTETLRLGLEENKPNLITRPFQWTHDDIQTVNMMDIEN 286
 QY 256 DKEPKGYRNAVEFEKKEIKVKTWVPEK 284
 DB 287 GTAPEA---AANWVANNPEKNEWIGEK 312
 RESULT 5
 ID Q926L5 LISIN PRELIMINARY; PRT; 289 AA.
 AC Q926L5;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Pseudogene, similar to glycine-betaine binding (ABC transporter).
 GN OrderedLocusNames=pl10045;
 OS Listeria innocua.
 OG Plasmid pl1100.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OC NCBI_TaxID=1642;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RX MEDLINE=21537279; PubMed=11679669; DOI=10.1126/science.1063447;
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Blocker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetoui F., Couve E., de Daruvar A., Denoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fshi H., Garcia-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
 RT "Comparative genomics of *Listeria species*,"
 RL Science 294:849-852(2001).
 DR EMBL; AL592102; CAC42043.1; -; Genomic_DNA.
 DR GO; GO:0005488; F:binding; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007210; OpuAC_ABC.
 DR Pfam; PF04069; OpuAC; 1.
 KW Complete proteome; Plasmid.
 SQ SEQUENCE 289 AA; 31737 MW; FASD1279BB86508F CRC64;
 Query Match 32.4%; Score 494.5; DB 2; Length 289;
 Best Local Similarity 37.9%; Pred. No. 1.7e-29;
 Matches 106; Conservative 54; Mismatches 107; Indels 13; Gaps 5;
 QY 9 IIFNPLSCDEKSKNLKS-----VKIGYVWGGTAATNVLKVPKMGYNAEIFSVTTS 64
 DB 15 LLFTFGSEDSYKTTVGADSGGQTVTLATVNWSEBIASNTNLAQVLKSAAGFNQITVDPA 74
 QY 65 IMYOVLASGKIDGTVSSVPTADKFFYKELTKFKFDLGANYEGTIQGFVPSVY-PISSI 123
 DB 75 IMFSSVAGQGLDAMVGGVPTTHQAYAEKYGDSVVDLGANLEGAISALTVPYTMEDINSI 134
 QY 124 SELKGGKPKFNKMGIDAGAGTQIVTEQALNYGLSKYELVPSSESVMLASLSSIKR 183
 DB 135 TDLTDEND---STTATEPGAGVNTSQAQNAKEVDNLSDWEVSSTGAMTAEILDQANN 191
 QY 184 NEWILVPLKPHWAFSRDYDIKFLDPLIMGIESVHTLVRLGLENDDFDAYVDFHFW 243
 DB 192 EEDIVVVGKPHWFMFMDYDLKMLDPPENVFGYEEIHSYAREGLKEDNPAYKIIDNFW 251
 QY 244 SDDLILPLMDK--NDKRGKEYRNAVEFEKKEIKVKTWV 281
 DB 252 EVEDMSSVMEELATDVEPEEADN---WIEANREIVDGLW 288
 RESULT 6
 Q9A6X6 BACTN PRELIMINARY; PRT; 277 AA.
 ID Q9A6X6 BACTN PRELIMINARY; PRT; 277 AA.
 AC Q9A6X6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Glycine betaine-binding protein.
 GN OrderedLocusNames=BT1749;
 OS Bacteroides thetaiotaomicron.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OC NCBI_TaxID=818;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=VPI-5482 / ATCC 29148;
 RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
 RA Chiang H.C., Hooper L.V., Gordon J.I.,
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis,"
 RL Science 299:2074-2076(2003).
 DR EMBL; AE016933; AA076856.1; -; Genomic_DNA.
 DR GO; GO:0005488; F:binding; IEA.
 DR GO; GO:0005215; F:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR007210; OpuAC_ABC.
 DR Pfam; PF04069; OpuAC; 1.
 KW Complete proteome.
 SQ SEQUENCE 277 AA; 31056 MW; B020C694CE12A21A CRC64;
 Query Match 31.3%; Score 478.5; DB 2; Length 277;
 Best Local Similarity 34.9%; Pred. No. 2.7e-28;
 Matches 98; Conservative 59; Mismatches 119; Indels 5; Gaps 2;


```
QY 244 SDDL--ILPLMDKNDKPKGKYRNNAVFEVKNKEIVKTVW 281
DB 243 ADNMNEVLMYAKDGMPEEA---AEKWVNAQDKVDEWL 279

RESULT 9
Q891M9_CLOTE PRELIMINARY; PRT; 298 AA.
AC Q891M9;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Glycine betaine-binding protein.
GN OrderedLocusNames=CTC02340;
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;
RA Bruggemann H., Baumer S., Fricke W.F., Wieser A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).
DR EMBL; AE015944; AAO36816.1; -; Genomic DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 298 AA; 33313 MW; A0E7CA5290977443 CRC64;

Query Match 30.2%; Score 460.5; DB 2; Length 298;
Best Local Similarity 37.6%; Pred. No. 76-27;
Matches 103; Conservative 44; Mismatches 114; Indels 13; Gaps 4;

QY 16 CDEKSSKNLKVIGYVNWGGTAAATNLVKVFF-EKMGYNAEIFSVTTSIMYQYLASGK 74
DB 30 CGKGAANKGKGVNLGYVNWAGVAMTNLAKVALEKMGYDELWTGEGAMFTSLSDGN 89
QY 75 IDGTSSWPTADKFFYEKLKTKFVDLGANYEGTIOGVVPSYVPISSISELKGKDGKPK 134
DB 90 IDAFLDANLFTVHDVTVTKYKDKIEDLGYNENARIGLVVPKNSDINSIEDLNKVKGKLG 149
QY 135 NKMGIGDAGAGTQIVTEQALNYGLSKYELVPSSEVMSLASLDSIKENEWILVPLMKP 194
DB 150 GKIIGDAGAGINSATERAKQYNL--EYELGSGGPTVTVMLEKAIDKKQDIVVTGWKP 207
QY 195 HWAFSYDIFKFLDDPLIMGGTIESVHTLVRLGLENDPDFAYVDFHFYMSDDLILPLM-- 252
DB 208 HWKFAWDLKFLEDPKKSIGBAENIHTYKGFEDKPMPEVAELFKNLNDSEQLGTLMGD 267
QY 253 ---DRNDKPKGKYRNNAVFEVKNKEIVKTVWPE 283
DB 268 IADSKKDPB-----EVAKEWMAKNEELVNSWIPK 296

RESULT 10
Q830X6_ENTFA PRELIMINARY; PRT; 569 AA.
AC Q830X6;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE Glycine betaine/L-proline ABC transporter, glycine betaine/L-proline-
DE binding/permease protein.
GN OrderedLocusNames=EF2642;
OS Enterococcus faecalis (Streptococcus faecalis).
```

```
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,
RA Taghert S.C., DeBoy R.T., Durkin S.A., Kolonay J.F., Madupu R.,
RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,
RA Khouri H.M., Utterback T.R., Radune D., Ketchum K.A., Dougherty B.A.,
RA Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074 (2003).
DR EMBL; AE016955; AAO82350.1; -; Genomic DNA.
DR TIGR; EF2642; -;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0015450; F:protein translocase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR00515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR InterPro; IPR002208; SecY_ABC.
DR Pfam; PF00528; BPD_transp_1; 1.
DR Pfam; PF04069; OpuAC; 1.
DR PRINTS; PR00303; SECYTRNLCASE.
DR PROSITE; PS0928; ABC_TM1; 1.
KW Complete proteome.
SQ SEQUENCE 569 AA; 61772 MW; AD4F445DD6CB9C3B CRC64;

Query Match 30.1%; Score 459; DB 2; Length 569;
Best Local Similarity 35.5%; Pred. No. 2e-26;
Matches 93; Conservative 59; Mismatches 102; Indels 8; Gaps 4;

QY 21 SSKNLKSVKIGYVNWGGTAAATNLVKVFFKMGYNAEIFSVTTSIMYQYLASGKIDGTVS 80
DB 312 TTNDKQISLSYGVDETVASTHVAEVLKMGYDKVTTPLDNAIMWSVAKGETDAMVG 371
QY 81 SWPFTADFFYKELKTKFVDLGANYEGTIOGVVPSYVPISSISELKGKDGKFKNMIGI 140
DB 372 AWPPTHAEQYKQYKDKLDDGENLKGAKLGLVPSYMDVDSIEDLSQAGK---KITGI 428
QY 141 DAGAGTQIVTEQALNYGLSKYELVPSSEVMSLASLDSIKENEWILVPLMKHPAFSR 200
DB 429 EPGAGVAAAEKTKAEYPNLKDWSVETSSSGAMTVAGLQAKNEDIVITGWSPHWPAK 488
QY 201 YDIKFLDDPLIMGGTIESVHTLVRLGLENDPDFAYVDFHFYMSDDLILPLMDKNDKPK 259
DB 489 YDLKYLADPKGTMGGEAHTMARQGLKEDQPEAYKVLDFNFWHTTKDMESVMLEINE--- 545
QY 260 GKEYRNNAV-EFVEKNKEIVKTVW 280
DB 546 GKDPQEAARDWVDHSHKQVAEW 567

RESULT 11
Q9KIF6_9LACT PRELIMINARY; PRT; 573 AA.
AC Q9KIF6;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE OpuABC.
GN Name=OpuABC;
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
```

RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20318987; PubMed=10860977; DOI=10.1073/pnas.97.13.7102;
RA van der Heide T., Poolman B.;
RT "Osmoregulated ABC-transport system of Lactococcus lactis senses water
stress via changes in the physical state of the membrane.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:7102-7106(2000).
DR EMBL; AF234619; AAF37879.1; -, Genomic_DNA.
DR PIR; C86806; C86806.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0005450; F:protein translocase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR InterPro; IPR002208; SecY.
DR Pfam; PF00528; BPD transp_1; 1.
DR PRINTS; PR00303; SECYTRNLCASE.
DR PROSITE; PS0928; ABC_TM1; 1.
DR PROSITE; PS0928; ABC_TM1; 1.
SQ SEQUENCE 573 AA; 61952 MW; DFB59CEAE8462A0F CRC64;
Query Match 29.5%; Score 451; DB 2; Length 573;
Best Local Similarity 34.9%; Pred. No. 8.3e-26;
Matches 99; Conservative 64; Mismatches 66; Indels 22; Gaps 7;
QY 6 LFFIIFLSCDEKSSKNLKVIGYVNWGGTAATNVLKVFPEKMGYNAEIFSVTTSI 65
DB 302 LALIIGAFSGMSFGKTASD-KKVDLVYVNWDEVASINVLTQAMKEHGFVKTALDNAV 360
QY 66 MYQYLASGKIDGTSSWVPTADKFFYEKLTKFVD-LGANYEGTIQGFVPSYVPISSIS 124
DB 361 AQWTVANGQADGNVSAWLPNTHKTQWKY-GKSVDLLGNLKGAKVGFPVPSYVNVNSIE 419
QY 125 ELKGGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSSEYMLASLDSISIKRN 184
DB 420 DLTNQANK---TITGIEPGAGVMAASEKTLSYDNLKDWKLVPPSSSGAMTVAGAIKQH 476
QY 185 EWILVPLKPHWAFSRDYIKFLDDPDLMGGIESVHTLVRLGLENDDFDYYVDFHFW- 243
DB 477 KDIVITGSPHWMFNKYDLKADPKGTMTGTSININTIVRKLKKENPEAYKVLDFKNWT 536
QY 244 SDDLILPLMD-KNDKEPGKEYRNAVEFVEKKEIIVKTWVPEKYK 286
DB 537 TKDMEAVMLDIQNGKTP-----EAAKNWIKDHQK 566
RESULT 12
Q7DAU8 LACLA PRELIMINARY; PRT; 573 AA.
AC Q7DAU8;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Name=busAB; OrderedLocNames=IL1451;
GN Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.CR-1697R;
RA Bolotin A., Winkler P., Mager S., Jallion O., Malarie K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006375; AA05549.1; -, Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0015450; F:protein translocase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0009306; P:protein secretion; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR InterPro; IPR002208; SecY.
DR Pfam; PF00528; BPD transp_1; 1.
DR PRINTS; PR00303; SECYTRNLCASE.
DR PROSITE; PS0928; ABC_TM1; 1.
DR PROSITE; PS0928; ABC_TM1; 1.
KW Complete proteome.
SQ SEQUENCE 573 AA; 61952 MW; DFB59CEAE8462A0F CRC64;
Query Match 29.5%; Score 451; DB 2; Length 573;
Best Local Similarity 34.9%; Pred. No. 8.3e-26;
Matches 99; Conservative 64; Mismatches 99; Indels 22; Gaps 7;
QY 6 LFFIIFLSCDEKSSKNLKVIGYVNWGGTAATNVLKVFPEKMGYNAEIFSVTTSI 65
DB 302 LALIIGAFSGMSFGKTASD-KKVDLVYVNWDEVASINVLTQAMKEHGFVKTALDNAV 360
QY 66 MYQYLASGKIDGTSSWVPTADKFFYEKLTKFVD-LGANYEGTIQGFVPSYVPISSIS 124
DB 361 AQWTVANGQADGNVSAWLPNTHKTQWKY-GKSVDLLGNLKGAKVGFPVPSYVNVNSIE 419
QY 125 ELKGGDKFKNMIGIDAGAGTQIVTEQALNYGLSKEYELVPSSSEYMLASLDSISIKRN 184
DB 420 DLTNQANK---TITGIEPGAGVMAASEKTLSYDNLKDWKLVPPSSSGAMTVAGAIKQH 476
QY 185 EWILVPLKPHWAFSRDYIKFLDDPDLMGGIESVHTLVRLGLENDDFDYYVDFHFW- 243
DB 477 KDIVITGSPHWMFNKYDLKADPKGTMTGTSININTIVRKLKKENPEAYKVLDFKNWT 536
QY 244 SDDLILPLMD-KNDKEPGKEYRNAVEFVEKKEIIVKTWVPEKYK 286
DB 537 TKDMEAVMLDIQNGKTP-----EAAKNWIKDHQK 566
RESULT 13
Q8EKX8 OCEIH PRELIMINARY; PRT; 297 AA.
AC Q8EKX8;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Glycine betaine ABC transporter glycine betaine-binding protein.
GN OrderedLocNames=OB3454;
OS Oceanobacillus theysensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=HT831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=1223376; DOI=10.1093/nar/gkf526;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theysensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; BA000028; BAC15410.1; -, Genomic_DNA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0030001; P:metal ion transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR006121; HeavyMe transp.
DR InterPro; IPR007210; OpuAC_ABC.
DR Pfam; PF04069; OpuAC; 1.
KW Complete proteome.
SQ SEQUENCE 297 AA; 32068 MW; C01B0FAC9BDCBE1A CRC64;

Query Match	29.3%;	Score 447;	DB 2;	Length 297;
Best Local Similarity	34.1%;	Pred. No. 7.5e-26;		
Matches	92;	Conservative 57;	Mismatches 109;	Indels 12; Gaps 3;
QY	18	EKKSSKNLKSIVGYNMGGETAATNLVKVPEKMGYNAPSPVTTTSIMYOVLASGKIDG 77		
DB	34	EGDSAENKEIELAYVEMDTEVASTHVIKGLVDLGYVELTPLDNLAIWAEVANGADG 93		
QY	78	TVSSWVPTADKPYEKLTKYFVDLGNAYEGTTQGFVPSYVPISSISELKGKDKFKNNM 137		
DB	94	MVAALPNTHAEQYESYGDQVESLGNLFGAKIGLVPEYMDVNSIEDL---SDRAGQTI 150		
QY	138	IGIDAGAGQIVTEQALNYGSLKEYELVPSSSEVMKSLDSSIKENWILVPLWKPWHA 197		
DB	151	TGLEPGAGVVAASEDAVETYNLDGMSQVTSNGAMATLGSAINDEBIIIVTGSHPWK 210		
QY	198	FSKYDILKFLDDPDLIMGGIESVHTLVRGLNDDFDAYVDFHFWYS-----DDLILPLMD 253		
DB	211	FQAYDLKYLEDEPGVGDATETIETVWREGLEDMEATYLDNFQDFAAMEEVMLNISE 270		
QY	254	KNDKEPGKRYNAVEFVKNEIKVKTWPE 283		
DB	271	GADPE-----QAAADVAEANKQKVDWTGD 295		
RESULT 14				
Q8K8R4 STRP3				
ID	Q8K8R4 STRP3 PRELIMINARY;	PRT;	561 AA.	
AC	Q8K8R4;			
DT	01-OCT-2002 (TremBLrel. 22, Created)			
DT	01-OCT-2002 (TremBLrel. 22, Last sequence update)			
DT	01-MAR-2004 (TremBLrel. 26, Last annotation update)			
DE	Putative glycine-betaine binding permease protein.			
GN	Name=opuABC; OrderedLocusNames=Spym3_0144;			
OS	Streptococcus pyogenes (serotype M3)			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=301448;			
RN	[1]			
RC	NUCLEOTIDE SEQUENCE.			
RP	STRAIN=MGAS315 / Serotype M3;			
RX	MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;			
RA	Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,			
RA	Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,			
RA	Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,			
RA	Schlievert P.M., Musser J.M.;			
RT	"Genome sequence of a serotype M3 strain of group A Streptococcus:			
RT	phage-encoded toxins, the high-virulence phenotype, and clone			
RT	emergence."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).			
RL	EMBL; AB014139; AAM78751.1; -; Genomic DNA.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005488; F:binding; IEA.			
DR	GO; GO:0015450; F:protein translocase activity; IEA.			
DR	GO; GO:0005215; F:transporter activity; IEA.			
DR	GO; GO:0009306; P:protein secretion; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	InterPro; IPR000515; BPD.transp.			
DR	InterPro; IPR007210; OpuAC.ABC.			
DR	InterPro; IPR002208; SecY.			
DR	Pfam; PF00528; BPD.transp_1; 1.			
DR	Pfam; PF04069; OpuAC; 1.			
DR	PRINTS; PR00303; SECYTRNLCASE.			
DR	PROSITE; PSS00928; ABC.TMI; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 561 AA; 61568 MW; 59BAE2C139AA55C2 CRC64;			
Query Match	29.0%;	Score 442.5;	DB 2;	Length 561;
Best Local Similarity	35.8%;	Pred. No. 3.6e-25;		
Matches	92;	Conservative 55;	Mismatches 103;	Indels 7; Gaps 5;
QY	26	KSVKIGYVNWGGETAATNLVKVPEKMGYNAPSPVTTTSIMYOVLASGKIDGTSWVPT 85		
DB	322	ETVNIAYVQDSEVASTHVIKGLVHTLPLDNVMTQVANGADFTSAMLV 391		
QY	86	ADKPYEKLTKFVDLGNAYEGTTQGFVPSYV-SSISELKGKDKFKNNMIGIDAGA 144		

Query Match	29.3%;	Score 447;	DB 2;	Length 297;
Best Local Similarity	34.1%;	Pred. No. 7.5e-26;		
Matches	92;	Conservative 57;	Mismatches 109;	Indels 12; Gaps 3;
QY	18	EKKSSKNLKSIVGYNMGGETAATNLVKVPEKMGYNAPSPVTTTSIMYOVLASGKIDG 77		
DB	34	EGDSAENKEIELAYVEMDTEVASTHVIKGLVDLGYVELTPLDNLAIWAEVANGADG 93		
QY	78	TVSSWVPTADKPYEKLTKYFVDLGNAYEGTTQGFVPSYVPISSISELKGKDKFKNNM 137		
DB	94	MVAALPNTHAEQYESYGDQVESLGNLFGAKIGLVPEYMDVNSIEDL---SDRAGQTI 150		
QY	138	IGIDAGAGQIVTEQALNYGSLKEYELVPSSSEVMKSLDSSIKENWILVPLWKPWHA 197		
DB	151	TGLEPGAGVVAASEDAVETYNLDGMSQVTSNGAMATLGSAINDEBIIIVTGSHPWK 210		
QY	198	FSKYDILKFLDDPDLIMGGIESVHTLVRGLNDDFDAYVDFHFWYS-----DDLILPLMD 253		
DB	211	FQAYDLKYLEDEPGVGDATETIETVWREGLEDMEATYLDNFQDFAAMEEVMLNISE 270		
QY	254	KNDKEPGKRYNAVEFVKNEIKVKTWPE 283		
DB	271	GADPE-----QAAADVAEANKQKVDWTGD 295		
RESULT 14				
Q8K8R4 STRP3				
ID	Q8K8R4 STRP3 PRELIMINARY;	PRT;	561 AA.	
AC	Q8K8R4;			
DT	01-OCT-2002 (TremBLrel. 22, Created)			
DT	01-OCT-2002 (TremBLrel. 22, Last sequence update)			
DT	01-MAR-2004 (TremBLrel. 26, Last annotation update)			
DE	Putative glycine-betaine binding permease protein.			
GN	Name=opuABC; OrderedLocusNames=Spym3_0144;			
OS	Streptococcus pyogenes (serotype M3)			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=301448;			
RN	[1]			
RC	NUCLEOTIDE SEQUENCE.			
RP	STRAIN=MGAS315 / Serotype M3;			
RX	MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;			
RA	Beres S.B., Sylva G.L., Barbian K.D., Lei B., Hoff J.S.,			
RA	Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,			
RA	Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,			
RA	Schlievert P.M., Musser J.M.;			
RT	"Genome sequence of a serotype M3 strain of group A Streptococcus:			
RT	phage-encoded toxins, the high-virulence phenotype, and clone			
RT	emergence."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).			
RL	EMBL; AB014139; AAM78751.1; -; Genomic DNA.			
DR	GO; GO:0016021; C:integral to membrane; IEA.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005488; F:binding; IEA.			
DR	GO; GO:0015450; F:protein translocase activity; IEA.			
DR	GO; GO:0005215; F:transporter activity; IEA.			
DR	GO; GO:0009306; P:protein secretion; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	InterPro; IPR000515; BPD.transp.			
DR	InterPro; IPR007210; OpuAC.ABC.			
DR	InterPro; IPR002208; SecY.			
DR	Pfam; PF00528; BPD.transp_1; 1.			
DR	Pfam; PF04069; OpuAC; 1.			
DR	PRINTS; PR00303; SECYTRNLCASE.			
DR	PROSITE; PSS00928; ABC.TMI; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 561 AA; 61568 MW; 59BAE2C139AA55C2 CRC64;			
Query Match	29.0%;	Score 442.5;	DB 2;	Length 561;
Best Local Similarity	35.8%;	Pred. No. 3.6e-25;		
Matches	92;	Conservative 55;	Mismatches 103;	Indels 7; Gaps 5;
QY	26	KSVKIGYVNWGGETAATNLVKVPEKMGYNAPSPVTTTSIMYOVLASGKIDGTSWVPT 85		
DB	322	ETVNIAYQVQDSEVASTHVIKNEGHVHTLPLDNVAMVMTQVANGADFTSAMLV 391		
QY	86	ADKPYEKLTKFVDLGNAYEGTTQGFVPSYV-SSISELKGKDKFKNNMIGIDAGA 144		

Db	382	THGQQYQKYKSLDGLGNLKGTLGLAVPKYMTDVNSIEDLSKQADQ---KITGIEPGA	438
QY	145	GTQIVTEQALNTYGLSKEYELVPSSSEVWMLASLDSIKNNEWILVPLWKPHWAFSRYDIK	204
Db	439	GIMAAAKKTLKEYHNLSSWELVAASTGAMTTSLDQAIIKKKDPVVVTAMSPHMMFAKYDLK	498
QY	205	FLDDPDLIMGGIESVHTLVRGLGLENDDPDAYVYVEDHEFYMSD-DLJLPLMDKNDKEPGKEY	263
Db	499	YLDKPKETFGSTENINTIARKGLKKDLNVYKIIDKFHWTQKDMAAVMLDIN-KGMSPE-	556
QY	264	RNAVEFEKNEKEIVKTW	280
Db	557	AAAKKWEANKSKVSSW	573

Search completed: January 24, 2006, 19:56:44
Job time : 63.3148 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 45.5886 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-10
Perfect score: 1090
Sequence: 1 RKFPEDLNIDVTMVPLOKK.....QFINYCSTLSKXKVFVDDNIM 204

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseq1980s.*
2: geneseq1990s.*
3: geneseq2000s.*
4: geneseq2001s.*
5: geneseq2002s.*
6: geneseq2003as.*
7: geneseq2003bs.*
8: geneseq2004s.*
9: geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	100.0	204	ADU98751	Adu98751 Borrelia
2	1090	100.0	377	ADU98753	Adu98753 Borrelia
3	152.5	14.0	385	ABM70978	Abm70978 Staphyloc
4	151.5	13.9	385	ABJ18960	Abj18960 Pathogen
5	150	13.8	375	ABE40038	Aeb40038 L. pneumo
6	150	13.8	382	ABE36646	Aeb36646 L. pneumo
7	145	13.3	156	AAI20035	Aay20035 B. burgdo
8	145	13.3	176	AAI20034	Aay20034 B. burgdo
9	141.5	13.0	408	ABP27526	Abp27526 Streptoco
10	137.5	12.6	379	ABE54455	Abb54455 Lactococc
11	137	12.6	376	ADV87963	Adv87963 Streptoco
12	137	12.6	376	ADV79216	Adv79216 Streptoco
13	137	12.6	376	ADV81406	Adv81406 Streptoco
14	136	12.5	376	ABP27525	Abp27525 Streptoco
15	135	12.4	381	AEA49154	Aea49154 L. rhanno
16	134	12.3	376	ABU01854	Abu01854 S. pneumo
17	134	12.3	376	ADK46982	Adk46982 Streptoco
18	134	12.3	409	ADR95677	Adr95677 Novel S.
19	134	12.3	409	AEA59547	Aea59547 Streptoco
20	118.5	10.9	295	AAI37169	Aay37169 Protein i
21	114.5	10.5	385	ABB48803	Abb48803 Listeria
22	112	10.3	391	ABP38845	Abp38845 Staphyloc
23	111	10.2	393	ADC97174	Adc97174 E. faeciu
24	107.5	9.9	442	AAB93553	Aab93553 Human pro

25	107.5	9.9	442	8	ABM81907	Abm81907 Tumour-as
26	105.5	9.7	401	9	ABM95708	Abm95708 M. xanthu
27	105	9.6	424	7	ABO71949	AbO71949 Pseudomon
28	104.5	9.6	397	7	ADH88000	Adh88000 Enterococ
29	102.5	9.4	394	2	AAI34978	Aay34978 Chlamydia
30	100.5	9.2	377	6	ABM67734	Abm67734 Photorhab
31	99.5	9.1	409	4	AAU67750	Aau67750 Propionib
32	99.5	9.1	409	6	ABM64269	Abm64269 Propionib
33	99.5	9.1	429	6	ABM65004	Abm65004 Propionib
34	98	9.0	387	7	ABO65431	AbO65431 Klebsiell
35	97	8.9	548	8	ADSO8193	AdS08193 Staphyloc
36	96.5	8.9	382	7	ADF04923	AdF04923 Bacterial
37	95.5	8.8	442	4	ABM95029	Abm95029 Human pro
38	93	8.5	350	4	ABB62280	Abb62280 Drosophil
39	93	8.5	4987	2	AAR10834	Aar10834 Rianodin
40	91.5	8.4	605	6	ABU26712	AbU26712 Protein e
41	91	8.3	1103	6	ABU48942	AbU48942 Protein e
42	90	8.3	605	7	ADC00089	Adc00089 Enterohae
43	89	8.2	2024	8	ADP25444	Adp25444 Plasmodiu
44	88.5	8.1	288	2	AAI19835	Aay19835 B. burgdo
45	88.5	8.1	310	2	AAI19834	Aay19834 B. burgdo

ALIGNMENTS

RESULT 1
ADU98751
ID ADU98751 standard; protein; 204 AA.
AC ADU98751;
XX
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 10.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
XX
PN WO2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
(TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
DR WPI; 2004-834155/82.
XX
DR N-PSDB; ADU98750.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises
providing to the subject at least one Borrelia antigen or its fragment.
XX
PS Claim 27; SEQ ID NO 10; 121pp; English.
XX
CC The invention describes a method of immunizing a subject comprising
providing to the subject at least one Borrelia antigen or its fragment to
induce an immune response. Also described are: an isolated polynucleotide
comprising a sequence having at least 17 contiguous nucleotides in common
with a sequence not given in the specification; an isolated polypeptide
having at least 5 consecutive amino acids of the sequence not given in
the specification; a vaccine composition comprising at least one Borrelia
antigen or at least one polynucleotide encoding a Borrelia antigen;
screening for at least one test polypeptide or test polynucleotide
encoding a polypeptide for an ability to produce an immune response;
preparing a vaccine; vaccinating a subject; treating a subject infected

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
CC ; and detecting antibodies to a herpesvirus. The method is useful in
CC immunizing a subject against Borrelia burgdorferi infection. This is the
CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
CC useful in immunizing a subject against infection by a member of the
CC Borrelia species.
XX
SQ Sequence 204 AA;

Query Match 100.0%; Score 1090; DB 8; Length 204;
Best Local Similarity 100.0%; Pred. No. 1e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RKPPFDLNDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 60
Db 1 RKPPFDLNDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 60

Qy 61 SEKLWFCALCECLSENGYINYEITNFALKGHESHRNKLWELKPHLGLGLYAVSLFCNDK 120
Db 61 SEKLWFCALCECLSENGYINYEITNFALKGHESHRNKLWELKPHLGLGLYAVSLFCNDK 120

Qy 121 NNNVRALIRKTGSFVKANNHLVTPELLEDEFFVYHFIQGLGTIQGVSLRALRLRFYNE 180
Db 121 NNNVRALIRKTGSFVKANNHLVTPELLEDEFFVYHFIQGLGTIQGVSLRALRLRFYNE 180

Qy 181 KQFFQFINYCSTLSKKFVDDNIM 204
Db 181 KQFFQFINYCSTLSKKFVDDNIM 204

RESULT 2
ADU98753
ID ADU98753 standard; protein; 377 AA.
XX AC ADU98753;
XX AC
XX 24-FEB-2005 (first entry)
XX DT
XX DE Borrelia burgdorferi antigenic polypeptide seqid 12.
XX KW antibacterial; vaccine; immune stimulation; immunity; antigen;
XX KW DNA library.
XX OS Borrelia burgdorferi.
XX XX
XX PN WO2004103269-A2.
XX PD 02-DEC-2004.
XX PF 17-OCT-2003; 2003WO-US033056.
XX PR 18-OCT-2002; 2002US-0419401P.
XX PA (MACR-) MACROGENICS INC.
XX PA (TEXA) UNIV TEXAS SYSTEM.
XX PI Sykes KF, Hale KS, Johnston SA;
XX WPI; 2004-834155/82.
XX DR N-PSDB; ADU98752.
XX XX
XX Immunizing a subject against Borrelia burgdorferi infection comprises
XX providing to the subject at least one Borrelia antigen or its fragment.
XX
XX Claim 27; SEQ ID NO 12; 121pp; English.
XX
XX The invention describes a method of immunizing a subject comprising
XX providing to the subject at least one Borrelia antigen or its fragment to
XX induce an immune response. Also described are: an isolated polynucleotide
XX comprising a sequence having at least 17 contiguous nucleotides in common
XX with a sequence not given in the specification; an isolated polypeptide
XX having at least 5 consecutive amino acids of the sequence not given in
XX the specification; a vaccine composition comprising at least one Borrelia

CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide
CC encoding a polypeptide for an ability to produce an immune response;
CC preparing a vaccine; vaccinating a subject; treating a subject infected
CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
CC ; and detecting antibodies to a herpesvirus. The method is useful in
CC immunizing a subject against Borrelia burgdorferi infection. This is the
CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
CC useful in immunizing a subject against infection by a member of the
CC Borrelia species.
XX
SQ Sequence 377 AA;

Query Match 100.0%; Score 1090; DB 8; Length 377;
Best Local Similarity 100.0%; Pred. No. 2.3e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RKPPFDLNDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 60
Db 141 RKPPFDLNDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 200

Qy 61 SEKLWFCALCECLSENGYINYEITNFALKGHESHRNKLWELKPHLGLGLYAVSLFCNDK 120
Db 201 SEKLWFCALCECLSENGYINYEITNFALKGHESHRNKLWELKPHLGLGLYAVSLFCNDK 260

Qy 121 NNNVRALIRKTGSFVKANNHLVTPELLEDEFFVYHFIQGLGTIQGVSLRALRLRFYNE 180
Db 261 NNNVRALIRKTGSFVKANNHLVTPELLEDEFFVYHFIQGLGTIQGVSLRALRLRFYNE 320

Qy 181 KQFFQFINYCSTLSKKFVDDNIM 204
Db 321 KQFFQFINYCSTLSKKFVDDNIM 344

RESULT 3
ABM70978
ID ABM70978 standard; protein; 385 AA.
XX AC ABM70978;
XX XX
XX 20-NOV-2003 (first entry)
XX DT
XX DE Staphylococcus aureus protein #218.
XX OS
XX KW Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;
XX KW enzymatic assay; antibiotic target.
XX OS Staphylococcus aureus.
XX PN WO200294868-A2.
XX PD 28-NOV-2002.
XX PF 27-MAR-2002; 2002WO-IB002637.
XX PR 27-MAR-2001; 2001GB-00007661.
XX PA (CHIR-) CHIRON SPA.
XX PI Masignani V, Mora M, Scarselli M;
XX WPI; 2003-120786/11.
XX DR N-PSDB; ACF72538.
XX XX
XX New Staphylococcus aureus protein, useful as a vaccine for treating or
XX preventing Staphylococcal infection, specifically an infection caused by
XX S. aureus, e.g. sepsis.
XX Claim 1; SEQ ID NO 436; 49pp; English.
XX
XX The invention relates to novel genes and encoded proteins from
XX Staphylococcus aureus. A composition comprising the S. aureus protein, a
XX nucleic acid encoding the protein, or an antibody to the protein, is

useful as a pharmaceutical, particularly as a vaccine for treating or preventing infection due to *Staphylococcus* bacteria, specifically an infection caused by *S. aureus*. The composition is particularly useful for treating or preventing sepsis in a patient. The composition can also be used for diagnostics. The protein is also used in an assay for enzymatic studies and as a target for antibiotics. This sequence represents one of the novel *S. aureus* proteins of the invention

Query Match 14.0%; Score 152.5; DB 6; Length 385;
Best Local Similarity 26.3%; Pred. No. 3.9e-07;
Matches 56; Conservative 44; Mismatches 94; Indels 29; Gaps 9;

Qy 7 LNIDMTVNWPLQKSHKRDLDKLLSYNPEHICFSDPICE-----EAGFV-LRDFD 56
Db 161 ISLDMLVHLPKQTIEDFEQSLDALDMDIQIHSSVGLILEPKTFQYNNYRKGLLEKLANED 220
Qy 57 NSIDSEKLMFCALECLSNQYINTEITNFALKGHESHKLNWELKPHLGLGLYA----- 111
Db 221 LGADMYQL---LMSKIQSPPHQYEISNFALDGHSEHNKYWFNERYYGAGASGYVD 277
Qy 112 -VSLFCNDKNNVRALIRKTSFVKANNHLVTFELLEDLLEFFVYHFTQGLGTIGQVSLR 170
Db 278 GVRTYNINPVNHYIKANKSKAILVSNKPSLT-ERMEE-EMFL-----GLRLNEGVSSS 330
Qy 171 ALRLRPEYN-EKQFFQFINTCYSTLSKRFVFDN 202
Db 331 RPKKKPQSIESVFGQTIN--NLKEKELIVEKN 361

RESULT 4
ABJ18960
ID ABJ18960 standard; protein; 385 AA.
XX ABJ18960;
AC
AC
DT 06-MAR-2003 (first entry)
XX
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 106.
XX
XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW autoimmune disease; HIV; hepatitis.
XX
OS Staphylococcus sp.
XX
XX WO200259148-A2.
XX
XX 01-AUG-2002.
XX
XX 21-JAN-2002; 2002WO-EP000546.
XX
XX 26-JAN-2001; 2001AT-00000130.
XX
XX (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
XX Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;
PI Minh DB, Vycvyska O, Etz H, Dryia A, Weichhart T, Hafner M;
PI Tempelmaier B;
XX
XX WPI; 2003-075410/07.
XX
XX Identifying, isolating and producing hyperimmune serum-reactive antigens
PT from a pathogen, for preparing vaccine or medicament for treating or
PT preventing e.g. staphylococcal infections, comprises providing antibody
PT preparation.
XX
XX Claim 21; Page 161; 252pp; English.
XX
XX The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC

CC	allergen, a tissue or host prone to auto-immunity, where the antigens are
CC	used in a vaccine, comprises providing antibody preparation from a plasma
CC	pool of a type of animal, or individual sera with antibodies against the
CC	specific pathogen, tumour, allergen, tissue or host prone to auto-
CC	immunity. The hyperimmune serum-reactive antigens comprising any of the
CC	62 sequences of 53-2261 amino acids fully defined in the specification,
CC	or their hyperimmune fragments are useful for the manufacture of a
CC	pharmaceutical preparation, particularly a vaccine against staphylococcal
CC	infections or colonisation against S. aureus or S. epidermidis. The
CC	preparation of antibodies is useful for the manufacture of a medicament
CC	for treating or preventing staphylococcal infections or colonisation
CC	against S. aureus or S. epidermidis. The antibody preparations may also
CC	be used for diagnostic and imaging purposes. Other conditions that can be
CC	treated include cancer, autoimmune diseases or infections caused by viral
CC	(e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This
CC	sequence represents a staphylococcal protein relating to the method for
CC	identifying and producing pathogen specific antigens of the invention
XX	
QQ	Sequence 385 AA;
	Query Match 13.9%; Score 151.5; DB 6; Length 385;
	Best Local Similarity 26.3%; Pred. No. 5e-07;
	Matches 56; Conservative 44; Mismatches 84; Indels 29; Gaps 9
QY	7 LNIDMTVMPLQKKSHLKRDLKELLSTYMEPHICFSDFICE-----BEGFV-LRDPD 56
DB	161 ISLDHYHLPKQTTFEDPEQSLLDALDMDIQHISSYGLILEPCTOFNNRYRKGLLKLPNED 220
QY	57 NSIDSEKLWFCALCLESNGVINYEITNFALKGHESRHKNLNWLKPHLGGLGYA---- 111
DB	221 LGADMYQL---LMSKIQSPHPQYEISNFALDGHESEHNKVWFNEEYFGAGASGYVD 277
QY	112 -VSLLFCNDKNNVRLIRTKTSFVKANNHLYVTLELLEDLFVFVHFIOGSLR 170
DB	278 GVRTYNINPVNYHIKAINKESKALIVSNKPSLT-ERWE-EWFL-----GLRLNEGVSSS 330
QY	171 ALRLRPYN-EKQFPQFINYCSTLSUKPFVFDN 202
DB	331 RFKKKFDQSIESVFGQTIN--NLKEKELIVEKN 361
RESULT 5	
AEB40038	ID AEB40038 standard; protein; 375 AA.
XX AC	AEB40038;
XX AC	
XX DT	08-SEP-2005 (first entry)
XX DE	L. pneumophila protein SEQ ID NO 4370.
XX KW	detection; infection; Antibacterial; Vaccine.
XX OS	Legionella pneumophila.
XX FN	WO2005049642-A2.
XX XX	
XX PD	02-JUN-2005.
XX PF	23-SEP-2004; 2004WO-IB003578.
XX PR	21-NOV-2003; 2003FR-00013687.
XX PA	(INSP) INST PASTEUR.
PA FA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
FA FA	(UPLY-) UNIV LYON 1 BERNARD CLAUDE.
PA FA	(CNRS) CNRS CENT NAT RECH SCI.
XX FI	Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;
PI FI	Ruenick C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;
FI FI	Jarraud S;
XX DR	WPI; 2005-388305/40.

XX New genome of Legionella pneumophila Paris strain and derived
PT polypeptides, useful for detection or identification of the strain and
PT for treatment and prevention of infections.
XX
PS Claim 3; SEQ ID NO 4370; 660pp; English.
XX
CC The invention relates to an isolated or purified nucleotide sequences (I)
CC from Legionella pneumophila Paris strain. (II), and their related
CC sequences or fragments, are useful as primers and probes for detection
CC and amplification, including differentiation between the Paris and
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant
CC (hybrid) polypeptides (II). (II) are also useful for preparation of
CC specific antibodies (Ab), also used for detection/identification of
CC Legionella, and some (I), specifically those involved in synthesis of
CC surface proteins, are targets for identification of inhibitors. (II), or
CC vectors that contain (I), are useful as vaccines and immunogenic
CC compositions, for treatment and prevention of infections by L.
CC pneumophila. The present sequence represents the amino acid sequence of a
CC L. pneumophila protein.
XX
SQ Sequence 375 AA;
Query Match 13.8%; Score 150; DB 9; Length 375;
Best Local Similarity 28.3%; Pred. No. 6.9e-07;
Matches 43; Conservative 21; Mismatches 40; Indels 48; Gaps 4;
Qy 1 RKFPD-LNIDMTVMPLQKSHLKRDLKELLSYMPHICF----- 40
Db 150 RKAGFDNLNLDIMHSLPNQSVTQGLQDLKTALSYQPEHLSWYQLTIBPNTVFKHTPPLP 209
Qy 41 ---SDFICEEGFVLDPDINSIDSEKLFWFCALCLESNGYINYEITNFALKGHESRHNKL 97
Db 210 SEEDYLLEEGFAL-----LHNSGNYRYEISAFSPKPEKQARHNIN 250
Qy 98 NWELKPHLGLGLYAVSLFLCNDKNNVRLAIR 129
Db 251 YWLFQDYLIGIGAGA-----HGKMTTPNAIR 276
RESULT 6
AEB36646
ID AEB36646 standard; protein; 382 AA.
XX
AC AEB36646;
XX
DT 08-SEP-2005 (first entry)
XX
DE L. pneumophila protein SEQ ID NO 978.
XX
KW detection; infection; Antibacterial; Vaccine.
XX
OS Legionella pneumophila.
XX
FN WO2005049642-A2.
XX
PD 02-JUN-2005.
XX
PF 23-SEP-2004; 2004WO-IB003578.
XX
PR 21-NOV-2003; 2003PR-00013687.
XX
PA (INSP) INST PASTEUR.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (UTLY-) UNIV LYON 1 BERNARD CLAUDE.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Buchrieser C, Tichit M, Etienne J, Ma L, Cazalet C, Glaser P;
PI Rusniok C, Bouchier C, Zidane N, Magnier A, Kunst F, Vandenesch F;
PI Jarraud S;
XX
DR WPI; 2005-388305/40.
XX

PT New genome of Legionella pneumophila Paris strain and derived
PT polypeptides, useful for detection or identification of the strain and
PT for treatment and prevention of infections.
XX
PS Claim 3; SEQ ID NO 978; 660pp; English.
XX
CC The invention relates to an isolated or purified nucleotide sequences (I)
CC from Legionella pneumophila Paris strain. (II), and their related
CC sequences or fragments, are useful as primers and probes for detection
CC and amplification, including differentiation between the Paris and
CC Philadelphia strains of Legionella pneumophila and to prepare recombinant
CC (hybrid) polypeptides (II). (II) are also useful for preparation of
CC specific antibodies (Ab), also used for detection/identification of
CC Legionella, and some (I), specifically those involved in synthesis of
CC surface proteins, are targets for identification of inhibitors. (II), or
CC vectors that contain (I), are useful as vaccines and immunogenic
CC compositions, for treatment and prevention of infections by L.
CC pneumophila. The present sequence represents the amino acid sequence of a
CC L. pneumophila protein.
XX
SQ Sequence 382 AA;
Query Match 13.8%; Score 150; DB 9; Length 382;
Best Local Similarity 28.3%; Pred. No. 7.1e-07;
Matches 43; Conservative 21; Mismatches 40; Indels 48; Gaps 4;
Qy 1 RKFPD-LNIDMTVMPLQKSHLKRDLKELLSYMPHICF----- 40
Db 157 RKAGFDNLNLDIMHSLPNQSVTQGLQDLKTALSYQPEHLSWYQLTIBPNTVFKHTPPLP 216
Qy 41 ---SDFICEEGFVLDPDINSIDSEKLFWFCALCLESNGYINYEITNFALKGHESRHNKL 97
Db 217 SEEDYLLEEGFAL-----LHNSGNYRYEISAFSPKPEKQARHNIN 257
Qy 98 NWELKPHLGLGLYAVSLFLCNDKNNVRLAIR 129
Db 258 YWLFQDYLIGIGAGA-----HGKMTTPNAIR 283
RESULT 7
AAAY20035
ID AAAY20035 standard; protein; 156 AA.
XX
AC AAAY20035;
XX
DT 19-JUL-1999 (first entry)
XX
DE B. burgdorferi antigenic protein, t229.aa.
XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX
OS Borrelia burgdorferi.
XX
FN WO9859071-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US012718.
XX
PR 20-JUN-1997; 97US-0050359P.
PR 22-JUL-1997; 97US-0053344P.
PR 22-JUL-1997; 97US-0053377P.
PR 03-SEP-1997; 97US-0057483P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX
DR WPI; 1999-189980/16.
DR N-ESDB; AAX61732.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the diagnosis, prevention and treatment of diseases caused
 XX by Borrelia, particularly Lyme disease.

XX Claim 12; Page 169; 275pp; English.

XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus

XX SQ Sequence 156 AA;

Query Match 13.3%; Score 145; DB 2; Length 156;
 Best Local Similarity 90.0%; Pred. No. 7.3e-07;
 Matches 27; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RKFPFDLNDMTVMPLQKKSHLKRDLKEL 30
 |||||
 Db 121 RKFPFDLNDMTVMPLQKKSHLKRDLQRI 150

RESULT 8

RAY20034
 ID AAY20034 standard; protein; 176 AA.

XX AC AAY20034;

XX DT 19-JUL-1999 (first entry)

XX DE B. burgdorferi antigenic protein, f229.aa.

XX KW Antigenic protein; vaccine; Lyme disease; infection; detection.

XX OS Borrelia burgdorferi.

XX PN W09859071-AL.

XX PD 30-DEC-1998.

XX PF 18-JUN-1998; 98WO-US012718.

XX PR 20-JUN-1997; 97US-0050359P.

XX PR 22-JUL-1997; 97US-0053344P.

XX PR 22-JUL-1997; 97US-0053377P.

XX PR 03-SEP-1997; 97US-0057483P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI (MEDI-) MEDIMMUNE INC.

XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

XX DR N-PSDB; AAX61731.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop
 XX products for the diagnosis, prevention and treatment of diseases caused
 XX by Borrelia, particularly Lyme disease.

XX Claim 12; Page 169; 275pp; English.
 XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides
 CC can be used in vaccines for eliciting protective antibodies to members of
 CC the Borrelia genus, particularly for the use against Lyme disease in
 CC humans and animals. They can be used for preventing or attenuating an
 CC infection caused by a member of the Borrelia genus. The products can also
 CC be used for detection of members of the Borrelia genus

XX SQ Sequence 176 AA;

Query Match 13.3%; Score 145; DB 2; Length 176;
 Best Local Similarity 90.0%; Pred. No. 8.6e-07;
 Matches 27; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RKFPFDLNDMTVMPLQKKSHLKRDLKEL 30
 |||||
 Db 141 RKFPFDLNDMTVMPLQKKSHLKRDLQRI 170

RESULT 9

ABP27526
 ID ABP27526 standard; protein; 408 AA.

XX AC ABP27526;

XX DT 02-JUL-2002 (first entry)

XX DE Streptococcus polypeptide SEQ ID NO 4228.

XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy.

XX OS Streptococcus pyogenes.

XX PN W0200234771-A2.

XX PD 02-MAY-2002.

XX PF 29-OCT-2001; 2001WO-GB004789.

XX PR 27-OCT-2000; 2000GB-00026333.

XX PR 24-NOV-2000; 2000GB-00028727.

XX PR 07-MAR-2001; 2001GB-00005640.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
 XX Tettelin H;

XX WPI; 2002-352536/38.

XX DR N-PSDB; ABN68157.

XX New Streptococcus protein for the treatment or prevention of infection or
 XX disease caused by Streptococcus bacteria, such as meningitis, and for
 XX detecting a compound that binds to the protein.

XX Claim 1; Page 3580; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN68044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins

XX SQ Sequence 408 AA;

Query Match 13.0%; Score 141.5; DB 5; Length 408;
 Best Local Similarity 25.1%; Pred. No. 6.1e-06;
 Matches 45; Conservative 32; Mismatches 91; Indels 11; Gaps 3;

Db 153 NISIDLIYALPGQTMDDVRSNVAKALSLNIPHLISLYSLILEHHTVPMNKMRRGKHLHLPTE 212

QY 58 SIDSEKLFWCALECLSENGVINYETITNFKLGHESRHNKLWELKPHLGLGLYAVSLFLC 117

Db 213 DLEAEFYEIISE-MERNGFHEYISNFTKPGFESRHNLMYWDNVEYGVGAGAGSYL-- 269

QY 118 NDKNNVRLIRKTGSF-----VKANNHLVTPELLEDEFFVYHFTIQGLGTIGVSLRA 171

Db 270 ----DGIR--YRNRGPIQHYLKGVSSEGNARLSSEVLSEKMEMEELFLGLRKEGVSIGK 323

QY 172 LRLRFEYN-EKQFFQFI 187

Db 324 FEQKFGTSFEKRYQIV 340

RESULT 12

ADV79216

ID ADV79216 standard; protein; 376 AA.

XX AC ADV79216;

XX DT 24-FEB-2005 (first entry)

XX DE Streptococcus agalactiae protein, SEQ ID 357.

XX KW Antibacterial; vaccine; bacterial infection.

XX OS Streptococcus agalactiae.

XX PN WO200292818-A2.

XX PD 21-NOV-2002.

XX PF 26-APR-2002; 2002WO-IB003059.

XX PR 26-APR-2001; 2001FR-00005642.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;

XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 357; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. The complete genome of Streptococcus agalactiae is given in ADV81204. Note: The present patent is an equivalent for the basic patent FR2824074A1, which contains only 2344 sequences.

XX SQ Sequence 376 AA;

Query Match 12.6%; Score 137; DB 8; Length 376;

Best Local Similarity 23.9%; Pred. No. 1.7e-05;

Matches 47; Conservative 36; Mismatches 90; Indels 24; Gaps 6;

QY 6 DLNIDMTVNMPLQKKSHLKRDLKELLSYMPHEIFCSDFICEEGFVLDRP-----DN 57

Db 153 NISIDLIYALPGQTMDDVRSNVAKALSLNIPHLISLYSLILEHHTVPMNKMRRGKHLHLPTE 212

QY 58 SIDSEKLFWCALECLSENGVINYETITNFKLGHESRHNKLWELKPHLGLGLYAVSLFLC 117

Db 213 DLEAEFYEIISE-MERNGFHEYISNFTKPGFESRHNLMYWDNVEYGVGAGAGSYL-- 269

QY 118 NDKNNVRLIRKTGSF-----VKANNHLVTPELLEDEFFVYHFTIQGLGTIGVSLRA 171

Db 270 ----DGIR--YRNRGPIQHYLKGVSSEGNARLSSEVLSEKMEMEELFLGLRKEGVSIGK 323

QY 172 LRLRFEYN-EKQFFQFI 187

Db 324 FEQKFGTSFEKRYQIV 340

RESULT 13

ADV81406

ID ADV81406 standard; protein; 376 AA.

XX AC ADV81406;

XX DT 24-FEB-2005 (first entry)

XX DE Streptococcus agalactiae protein, SEQ ID 2547.

XX KW Antibacterial; vaccine; bacterial infection.

XX OS Streptococcus agalactiae.

XX PN WO200292818-A2.

XX PD 21-NOV-2002.

XX PF 26-APR-2002; 2002WO-IB003059.

XX PR 26-APR-2001; 2001FR-00005642.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;

XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus agalactiae for the development of vaccines, diagnostic tools, DNA chips and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 2547; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The nucleotide sequences encode polypeptides of S. agalactiae involved in the synthesis of amino acids, cell membranes, intermediate (central) metabolism, energetic metabolism, fatty acid and phospholipid metabolism, nucleotide metabolism including purines, pyrimidines and/or nucleosides, regulatory functions, replication, transcription, translation, protein transport, adaptation to atypical conditions, sensitivity to medicines and/or analogues, functions related to transposons, biosynthesis of cofactors, prosthetic groups and transporters, cell membrane proteins and cellular machinery. (I) are useful for the detection and/or amplification of nucleic acids. Pharmaceutical composition comprising (I) or (II) are useful for treatment of a bacterial S. agalactiae infection. The complete genome of Streptococcus agalactiae is given in ADV81204. Note: The present patent is an equivalent for the basic patent FR2824074A1, which contains only 2344 sequences.

XX

SQ Sequence 376 AA;

Query Match 12.6%; Score 137; DB 8; Length 376;
Best Local Similarity 23.9%; Pred. No. 1.7e-05;
Matches 47; Conservative 36; Mismatches 90; Indels 24; Gaps 6;

QY 6 DLNIDMTVMPLQKSHLKRDLKELLSYMPHEICFSDPICEERGFVLDP-----DN 57
DB 153 NISIDLIYALPGQTMDDVRNVAKALSLNIPHLISLILIEHHTVFMKMRGKHLHLPTE 212

QY 58 SIDSEKLMFCALECLSNVYIEITNFALKGHESRHNKLNWELKPHLGLGLYAVSLFLC 117
DB 213 DLEAEMFEYIIE-MERNGFEHYEISNFTKPGFESRHNLMYWDNVEYGVGAGASGYL-- 269

QY 118 NDKNNVRALIRKTSF-----VKANNHLVTPLELLEDFVYHFIOGLGTIOGVSIRA 171
DB 270 ----DGIR--YRNRGPIQHYLKGVSNGARLSBEVLKSNEMBEELFLGLRKKEGVSIGK 323

QY 172 LRLRFEYN-EKQFFQFI 187
DB 324 FEQKFGTSPEKRYGQIV 340

RESULT 14
ABP27525
ID ABP27525 standard; protein; 376 AA.
XX AC ABP27525;
XX XX
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 4226.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX XX
XX FN WO200234771-A2.
XX XX
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX XX
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX XX
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX XX
XX PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
XX PI Tettelin H;
XX XX
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN68156.
XX XX
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX XX
XX FS Claim 1; Page 3580; 4525pp; English.
XX XX
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX XX
SQ Sequence 376 AA;

Query Match 12.5%; Score 136; DB 5; Length 376;
Best Local Similarity 23.9%; Pred. No. 2.1e-05;
Matches 47; Conservative 36; Mismatches 90; Indels 24; Gaps 6;

QY 6 DLNIDMTVMPLQKSHLKRDLKELLSYMPHEICFSDPICEERGFVLDP-----DN 57
DB 153 NISIDLIYALPGQTMDDVRNVAKALSLNIPHLISLILIEHHTVFMKMRGKHLHLPTE 212

QY 58 SIDSEKLMFCALECLSNVYIEITNFALKGHESRHNKLNWELKPHLGLGLYAVSLFLC 117
DB 213 DLEAEMFEYIIE-MERNGFEHYEISNFTKPGFESRHNLMYWDNVEYGVGAGASGYL-- 269

QY 118 NDKNNVRALIRKTSF-----VKANNHLVTPLELLEDFVYHFIOGLGTIOGVSIRA 171
DB 270 ----DGIR--YRNRGPIQHYLKGVSNGARLSBEVLKSNEMBEELFLGLRKKEGVSIGK 323

QY 172 LRLRFEYN-EKQFFQFI 187
DB 324 FEQKFGTSPEKRYGQIV 340

RESULT 15
AEA49154
ID AEA49154 standard; protein; 381 AA.
XX AC AEA49154;
XX XX
XX DT 25-AUG-2005 (first entry)
XX DE L. rhamnosus polypeptide #18.
XX KW Gene fusion; wound healing; immunostimulant; antilipemic; vulnerary.
XX OS Lactobacillus rhamnosus.
XX XX
XX PN WO2005056801-A1.
XX XX
XX PD 23-JUN-2005.
XX PF 15-DEC-2003; 2003WO-NZ000278.
XX XX
XX PR 15-DEC-2003; 2003WO-NZ000278.
XX XX
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PA (FONT-) FONTERRA COOP GROUP LTD.
XX XX
XX FI Glenn M, Havukkala I, Lubbers MW, Dekker J;
XX XX
XX DR WPI; 2005-435589/44.
XX DR N-PSDB; AEA49072.
XX XX
XX PT Isolated Lactobacillus polynucleotide, useful for modulating activity of
XX PT polypeptide in a cell, and for production of more effective probiotic
XX PT bacteria.
XX XX
XX FS Claim 18; SEQ ID NO 98; 323pp; English.
XX XX
XX CC The invention relates to an isolated Lactobacillus rhamnosus
XX CC polynucleotide and the polypeptide it encodes. The invention also relates
XX CC to a genetic construct comprising the polynucleotide, a transgenic host
XX CC cell comprising the genetic construct, a transgenic organism comprising
XX CC the host cell or its progeny and a method of modulating the activity of a

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 7.73918 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-10
Perfect score: 1090
Sequence: 1 RKPPFDLNDMTVMPLQKK.....QFINYCSTLSKKFVDDNIM 204
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	100.0	377	2 G70181	oxygen-independent
2	205	18.8	374	2 G97057	coproporphyrinogen
3	175.5	16.1	396	2 AD2425	coproporphyrinogen
4	151.5	13.9	374	2 F89939	hypothetical prote
5	148.5	13.6	370	2 H70423	oxygen-independent
6	141.5	13.0	385	2 G83817	coproporphyrinogen
7	138.5	12.7	374	2 D72288	hypothetical prote
8	137.5	12.6	379	2 C86767	oxygen-independent
9	136	12.5	393	2 D71354	probable oxygen-in
10	134.5	12.3	378	2 B71564	probable copropor
11	134	12.3	376	2 A98030	coproporphyrinogen
12	134	12.3	376	2 B95164	hypothetical prote
13	130.5	12.0	399	2 C71728	probable oxygen-in
14	126.5	11.6	412	1 S75358	coproporphyrinogen
15	124.5	11.4	366	1 B69640	coproporphyrinogen
16	124	11.4	391	2 G82320	probable oxygen-in
17	122.5	11.2	400	2 C97277	hypothetical prote
18	117	10.7	345	2 E71843	oxygen-independent
19	116	10.6	376	2 AD0116	probable oxygen-in
20	114.5	10.5	385	2 AD1259	coproporphyrinogen
21	114.5	10.5	385	2 AH1621	coproporphyrinogen
22	113	10.4	352	1 B64673	coproporphyrinogen
23	107	9.8	383	2 B82673	conserved hypothet
24	106	9.7	399	2 AE2616	hypothetical prote
25	106	9.7	402	2 E97398	probable oxygen-in
26	105.5	9.7	388	2 B73597	probable copropor
27	105	9.6	384	2 B83597	hypothetical prote
28	104.5	9.6	376	2 B84994	coproporphyrinogen
29	102.5	9.4	373	2 B86538	coproporphyrinogen

coproporphyrinogen
probable oxidoredu
probable oxygen-in
hypothetical prote
probable oxidase I
probable oxidase I
probable oxidase Y
hypothetical prote
coproporphyrinogen
flagellar hook-len
coproporphyrinogen
coproporphyrinogen
cyanodine receptor
RNA primase (EC 2.
hypothetical prote
multiple banded an
glycosyltransferas
probable membrane

ALIGNMENTS

RESULT 1

G70181

oxygen-independent coproporphyrinogen III oxidase homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: G70181

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White,

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: G70181

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-377 <KLE>

A;Cross-references: UNIPROT:O51600; UNIPARC:UPI0000057576; GB:AE001167; GB:AE000783; NID

A;Experimental source: strain B31

Query Match 100.0%; Score 1090; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 6.1e-88;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RKPPFDLNDMTVMPLQKKSHLRDLKELLSYMPHEICFSDPICBEEGFVLRDFDONSID 60
|||||
Db 141 RKPPFDLNDMTVMPLQKKSHLRDLKELLSYMPHEICFSDPICBEEGFVLRDFDONSID 200

QY 61 SEKLWFCALCELESNGYINYEITNFALKGHESHNNKLNWELKPHLGLGYAVSLLCNDK 120
|||||
Db 201 SEKLWFCALCELESNGYINYEITNFALKGHESHNNKLNWELKPHLGLGYAVSLLCNDK 260

QY 121 NNNVRALIRKTSFVKANNHLVTFELLEDLFPVYHFIOGLGTIOGVSRLALRLFEYNE 180
|||||
Db 261 NNNVRALIRKTSFVKANNHLVTFELLEDLFPVYHFIOGLGTIOGVSRLALRLFEYNE 320

QY 181 KQFFQFINYCSTLSKKFVDDNIM 204
|||||
Db 321 KQFFQFINYCSTLSKKFVDDNIM 344

RESULT 2

G97057

coproporphyrinogen III oxidase [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004

C;Accession: G97057

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A;Reference number: A96900; MUID:21359325; PMID:21359325

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
C;Accession: D71354
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <COL>
A;Cross-references: UNIPROT:O83216; UNIPARC:UPI00000D3217; GB:AE001202; GB:AE000520; NID
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0186
C;Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.5%; Score 136; DB 2; Length 393;
Best Local Similarity 25.6%; Pred. No. 0.00024;
Matches 50; Conservative 26; Mismatches 73; Indels 46; Gaps 8;

Qy 7 LNIDMTVMPLQKKSHLKRDLKELLSYMPHICFSDPIC-----EEEGFVLRFDFNSI 59
Db 168 ISADLIAGLRQATARMVREDIDELLSFGLRHVSLYG-LCVPHPTQTQERIA----- 218

Qy 60 DSEKMFCALECLSGNGYINYEITNPA-LKGHESRHNKLWELKPHLGLGLYAVSLPCN 118
Db 219 ---ALWAHGSAYLVRAGFNRYELSNFARTAADESAAHNRAYWRMAPHAGVGPAGVGTFRVN 275

Qy 119 DKNNN-----VRALIRKTGSFVKANNHL-----VTPELLEDLEFFVYH-----FIQGL 161
Db 276 LSLUSKEGAWAIRSVRK-----HLGQYLAECVRENVEYHEFLTEHMCVQALLMGL 326

Qy 162 GTIQGVSLRALRLRF 176
Db 327 RLEQGLDVVTFARP 341

RESULT 10
B71564
Probable coproporphyrinogen III oxidase - Chlamydia trachomatis (serotype D, strain UW3/
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: B71564
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:99000809; PMID:9784136
A;Accession: B71564
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <ARN>
A;Cross-references: UNIPROT:O84055; UNIPARC:UPI00000D32E8; GB:AE001279; GB:AE001273; NID
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: hemN 1
C;Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.3%; Score 134.5; DB 2; Length 378;
Best Local Similarity 27.9%; Pred. No. 0.00031;
Matches 43; Conservative 26; Mismatches 70; Indels 15; Gaps 4;

Qy 6 DLNIDMTVMPLQKKSHLKRDLKELLSYMPHICFSDPICEEGFVL-----RDFDNDSDS 61
Db 155 NVSADLIYGLTQTSIDFIVDLHQALSLPIQHSIYNLTIDPHTSFYKHKRILPSIADD 214

Qy 62 EKLMFCALECLSGNGYINYEITNPA-LKGHESRHNKLWELKPHLGLGLYAVSLPCN 118
Db 215 DSLAEMALABELLENQGFTRFELASAKNQAKSHNTYYTAKPFLGLGVSASQYL--- 271

Qy 119 DKNNNVA--LIRKTGSFVKANNHLVTPELLEDL 150
Db 272 ---HGIRSKNLSRISHYLRARAHQHLPTLESMEEL 302

RESULT 11

A98030
coproporphyrinogen oxidase (EC 1.3.3.3) [imported] - Streptococcus pneumoniae (strain R6
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: A98030
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: A98030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <KUR>
A;Cross-references: UNIPROT:Q97Q22; UNIPARC:Q8DPA6; UNIPROT:UPI0000051830; GB:AE007317;
C;Genetics:
A;Gene: hemN
C;Superfamily: oxygen-independent coproporphyrinogen oxidase
C;Keywords: oxidoreductase

Query Match 12.3%; Score 134; DB 2; Length 376;
Best Local Similarity 26.4%; Pred. No. 0.00034;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVMPLQKKSHLKRDLKELLSYMPHICFSDPICEEGFVLRFD----- 55
Db 151 FDNISIDLIALPGQTMQVKENVAKAIGLDIPHMSLYSLILENHTVPMNRRGKPLP 210

Qy 56 DNSIDSEKLMFCALECLSGNGYINYEITNPA-LKGHESRHNKLWELKPHLGLGLYAVSL 115
Db 211 KEELAEAFYIIAE-LERAGPEHYELSNFKPGFSRHNLMYWDNAEYIGIGAGASGYV 269

Qy 116 FCNDKNNVR-----ALIRKTGSFVKANNHLVTPELLEDLEFFVYHFIQGLGTICGSLRA 171
Db 270 -----NGVRYKNRGPTRHYLSAVEGNACITDHLISQKEQBEEMFLGLRKKSGVSM-- 321

Qy 172 LRLRFEYNEKQFF 184
Db 322 --ARFEKFGQSF 332

RESULT 12
B95164
hypothetical protein SPI409 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: B95164
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: B95164
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <KUR>
A;Cross-references: UNIPROT:Q97Q22; UNIPARC:Q8DPA6; UNIPARC:UPI0000051830; GB:AE005672;
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SPI409
C;Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.3%; Score 134; DB 2; Length 376;
Best Local Similarity 26.4%; Pred. No. 0.00034;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVMPLQKKSHLKRDLKELLSYMPHICFSDPICEEGFVLRFD----- 55
Db 151 FDNISIDLIALPGQTMQVKENVAKAIGLDIPHMSLYSLILENHTVPMNRRGKPLP 210

Qy 56 DNSIDSEKLFCALECSNGYINYEITNFALKGHESRHNKLNWELKPHLGILGYAVSLL 115
 Db 211 KEELAEAFYIIAE-LERAGFEHYEISNFKPGFESRHNLMYWDNAEYYIGAGAGSYV 269
 Qy 116 FCNDKNNVR---ALIRKTSFVKANNHLVTFELLEDLEFFVYHFIOGLGTIOGVSURA 171
 Db 270 -----NGVRKNGPPIHYLSAVEEGNACITEDHLSQKEQMEEMFGLRKKSGVSM-- 321
 Qy 172 LRLRFEYNEKQFF 184
 Db 322 --ARFEKFGQSF 332

RESULT 13
 C71728
 probable oxygen-independent coproporphyrinogen III oxidase (hemN) RP175 - Rickettsia pro
 C;Species: Rickettsia prowazekii
 C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
 C;Accession: C71728
 R;Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U
 Nature 396, 133-140, 1998
 A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A;Reference number: A71630; MUID:99039499; PMID:9823893
 A;Accession: C71728
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-399 <AND>
 A;Cross-references: UNIPROT:Q92DY8; UNIPARC:UPI00000D3775; GB:AJ235270; GB:AJ235269; NID
 A;Experimental source: strain Madrid E
 C;Genetics:
 A;Gene: hemN; RP175
 C;Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.0%; Score 130.5; DB 2; Length 399;
 Best Local Similarity 22.5%; Pred. No. 0.00074;
 Matches 43; Conservative 37; Mismatches 72; Indels 39; Gaps 6;
 Qy 24 KRDLEKLSYMPHICFSDFTCEBEGFVLRDPDN-----SIDSEKLFCALECSNG 76
 Db 189 EKELQAMYLATSHALYQLTIKGTTPPYKFKGNLILPTSDRAAEVYWNHYLESKK 248
 Qy 77 YINYEITNFALKGHESRHNKLNWELKPHLGILGYAVSLLFCNDKNNVRALI----- 128
 Db 249 YFRYEISNRYALGHECHLNLTWYNSYLGICPGAHSRII--ENSNSYLAITMLHNPQW 306
 Qy 129 -----RKTGSFVKANNHLVTFELLEDLEFFVYHFIOGLGTIOGVSLEALRLR----- 175
 Db 307 LDVSTKNVG--IQTHSKLTNKBIEEI-----LMMGLRLRGKIKSKLEQRLSTALEN 358
 Qy 176 -FEYNEKQFFQ 185
 Db 359 ILDINNQLHYQ 369

RESULT 14
 S75358
 coproporphyrinogen oxidase (BC 1.3.3.3) III, oxygen-independent - Synecocystis sp. (str
 C;Species: Synecocystis sp.
 A;Variety: PCC 6803
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C;Accession: S75358
 R;Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
 s.
 A;Reference number: S74322; MUID:97061201; PMID:8905231
 A;Accession: S75358
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-412 <KAN>
 A;Cross-references: UNIPROT:P73245; UNIPARC:UPI000013B4A6; EMBL:D90904; GB:AB001339; NID

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Genetics:
 A;Start codon: GTG
 C;Superfamily: oxygen-independent coproporphyrinogen oxidase
 C;Keywords: oxidoreductase

Query Match 11.6%; Score 126.5; DB 1; Length 412;
 Best Local Similarity 27.7%; Pred. No. 0.0017;
 Matches 33; Conservative 20; Mismatches 51; Indels 15; Gaps 2;
 Qy 8 NIDTMNMPLOKKSILKRLDKELLSYMPHICFSDFTCEBEGFVLRDPDNISIDSEKL--- 64
 Db 167 SLDLITGLPQTAADWHSSLTALAAGPKHISCDVLVLEPQTV---FDRWEQGRKLAVP 222
 Qy 65 -----WFCALCELESNGYINYEITNFALKGHESRHNKLNWELKPHLGILGYAVSLL 115
 Db 223 PPSADFYRHGQVLTQAGFHHYEISNYGRPGHQRHNQIYWRNLPPYGLGMSATSYI 281

RESULT 15
 B69640
 coproporphyrinogen oxidase (EC 1.3.3.3) III, oxygen-independent hemN - Bacillus subtilis
 C;Species: Bacillus subtilis
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C;Accession: B69640
 R;Kunst, F.; Ogasawara, N.; Mozer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mausel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A;Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: B69640
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-366 <KUN>
 A;Cross-references: UNIPROT:P54304; UNIPARC:UPI000006075C; GB:Z99117; GB:AL009126; NID:G
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: hemN
 C;Superfamily: oxygen-independent coproporphyrinogen oxidase
 C;Keywords: oxidoreductase

Query Match 11.4%; Score 124.5; DB 1; Length 366;
 Best Local Similarity 26.9%; Pred. No. 0.0023;
 Matches 50; Conservative 27; Mismatches 80; Indels 29; Gaps 7;
 Qy 1 RKFPFD-LNIDTMNMPLOKKSILKRLDKELLSYMPHICFSDFTCEBEGFVLRDPDN 59
 Db 148 REIGFENISLDLMFGLPGQTLKLEHSINTALSDEAHYSVYSLVPEKTV---FYNL 203
 Qy 60 DSEKL-----WFCALCELESNGYINYEITNFALKGHESRHNKLNWELKPHLG 108
 Db 204 QKGRLLHPQOEAEYVEIVMSKMEAHGHIHQYEISNFAKAGMSKHNLTYSNSEQYFG 263
 Qy 109 LYAVSLFLPCNDKNNVRA-----LIRKTSFVKANNHLVTFELLEDLEFFVYHFIOGL 162
 Db 264 AGAHGYI-GGTRTVNNGVFKHYIDLTAEGK-FPYRDTHVTEEQIBEEMFL-----GLR 316
 Qy 163 TIQGV 168
 Db 317 KTAVS 322

Search completed: January 24, 2006, 19:59:03
Job time : 8.73918 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 43.9352 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-10

Perfect score: 1090

Sequence: 1 RKPPFDLNDVTNMPLOKK.....QFINYCSTLSKKVFDNDNIM 204

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05_80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1090	100.0	377	051600 BORBU	051600 borrelia bu
2	934	85.7	377	Q66014 BORGA	Q66014 borrelia ga
3	210	19.3	365	Q8R695 FUSNN	Q8R695 fusbacteri
4	205.5	18.9	375	Q8RB71 THETN	Q8RB71 thermoanaer
5	205	18.8	374	Q97J25 CIOAB	Q97J25 clostridium
6	192.5	17.7	382	Q892Q7 CLOTE	Q892Q7 clostridium
7	190.5	17.5	165	Q7P440 FUSNV	Q7P440 fusbacteri
8	189	17.3	377	Q8X188 CLOPE	Q8X188 clostridium
9	175.5	16.1	396	Q8YMH7 ANASP	Q8YMH7 anabaena sp
10	175	16.1	408	Q7NC94 GIOVI	Q7NC94 gloeobacter
11	172.5	15.8	391	Q73MP2 TREDE	Q73MP2 treponema d
12	163.5	15.0	375	Q8KAY8 CHLTE	Q8KAY8 chlorobium
13	161.5	14.8	374	Q6GGB7 STAAR	Q6GGB7 staphylococ
14	157.5	14.4	374	Y498_BUCBP	Q89A47 buchnera ap
15	154	14.1	378	Q5LZC1 STRT1	Q5LZC1 streptococc
16	151.5	13.9	374	Q6G8V4 STAA5	Q6G8V4 staphylococ
17	151.5	13.9	374	Q7A0S0 STAAW	Q7A0S0 staphylococ
18	151.5	13.9	374	Q7ASC2 STAAW	Q7ASC2 staphylococ
19	151.5	13.9	374	Q99TR6 STAAW	Q99TR6 staphylococ
20	151.5	13.9	385	Q5HFH7 STAAC	Q5HFH7 staphylococ
21	151	13.9	378	Q5M3X3 STRT2	Q5M3X3 streptococc
22	150	13.8	374	Q9RPI0 9FIRM	Q9RPI0 desulfitoba
23	150	13.8	375	Q5X2T4 LEGPA	Q5X2T4 legionella
24	150	13.8	375	Q5ZT11 LEGPH	Q5ZT11 legionella
25	150	13.8	376	Q7MMW0 PORGI	Q7MMW0 porphyromon
26	149.5	13.7	371	Q7VIJ0 HELHP	Q7VIJ0 helicobacte
27	148.5	13.6	370	Q67418 AQUAE	Q67418 aquifex aeo
28	146	13.4	375	Q5WU92 LEGPL	Q5WU92 legionella
29	145.5	13.3	374	Q4FF16_9THEM	Q4FF16 thermotoga
30	145	13.3	408	Q7UTS9_PROMP	Q7UTS9 prochloroco
31	142.5	13.1	351	Q8EVQ8_MYCPE	Q8EVQ8 mycoplasma

32 142.5 13.1 383 2 05QY62 IDILO
33 141.5 13.0 376 2 099ZW6_STRPY
34 141.5 13.0 376 2 08PI77_STRP8
35 141.5 13.0 385 2 09KD75_BACHD
36 141.5 13.0 398 2 0878L1_STRP3
37 141.5 13.0 408 2 08K7R5_STRP3
38 141.5 13.0 408 2 05XCD8_STRP6
39 141 12.9 349 2 04HKR7_CAMLA
40 140 12.8 378 2 081LR9_BACAN
41 140 12.8 379 2 06HT74_BACAN
42 140 12.8 379 2 0634M4_BACZ
43 139.5 12.8 383 2 093R30_TETHA
44 139 12.8 379 2 05KWZ4_GEOKA
45 138.5 12.7 374 2 09X0P9_THEME

ALIGNMENTS

RESULT 1

ID 051600 BORBU PRELIMINARY; PRT; 377 AA.
AC 051600;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase, putative.
GN OrderedLocusNames=BB0656;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RE MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL Nature 390:580-586(1997).
DR EMBL; AE001167; AAC67009.1; -; Genomic_DNA.
DR PIR; G70181; G70181.
DR TIGR; BB0656; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR InterPro; IPR006638; ELP3/MiaB/NifB.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF04055; Radical_SAM; I.
DR SMART; SM00729; ELP3; I.
KW Complete proteome.
SQ SEQUENCE 377 AA; 44756 MW; B07859CFC7793561 CRC64;

Query Match 100.0%; Score 1090; DB 2; Length 377;
Best Local Similarity 100.0%; Pred. No. 3.9e-87;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RKPPFDLNDVTNMPLOKKSHLKRDLKELLSYMPHICFSDFCIEGFLVRFDNSID 60
Dy 141 RKPPFDLNDVTNMPLOKKSHLKRDLKELLSYMPHICFSDFCIEGFLVRFDNSID 200
Qy 61 SEKIWFCALECSNGYINYEITNFALKGHESRHNKLNWELKPHLGLGLVAVSLFCNDK 120
Dy 201 SEKIWFCALECSNGYINYEITNFALKGHESRHNKLNWELKPHLGLGLVAVSLFCNDK 260
Qy 121 NNVRVALIRKTSFVKANNHLVTFPELLEDLEFFVYHFIQGLTIGQVSLRALRPFYNE 180

```
Db 261 NNNVRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIOGLGTIGQVSLRALRLRFYNE 320
Qy 181 KQFFQFNYCSTLSKKFVFDNIM 204
Db 321 KQFFQFNYCSTLSKKFVFDNIM 344

RESULT 2
Q660L4_BORGA
ID Q660L4_BORGA PRELIMINARY; PRT; 377 AA.
AC Q660L4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE HemN-related protein.
GN OrderedLocusNames=BG0679;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romaldi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Suehnel J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000013; AAU07507.1; -; Genomic_DNA.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR Complete proteome.
KW SEQUENCE 377 AA; 44968 MW; C5192B7F0CD24051 CRC64;

Query Match 85.7%; Score 934; DB 2; Length 377;
Best Local Similarity 84.8%; Pred. No. 1.8e-73;
Matches 173; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

Qy 1 RKFPFDLNDMTVMPLQKSHLRDLKELLSYMPHEICFSDPICBEGFVLRFDSID 60
Db 141 RKFSFDLNDMTINIPQKSDLRDLKELLSYMPHEICFSDPICBEGFVLRFDSIGD 200

Qy 61 SEKLWFCALCELSNGYINVTITPALKGHESRNKLNWELKPHGLGLYAVSLFCNDK 120
Db 201 SEKLWFALEYLSNGYINVTITPMLKGHSKKNLNWELKPHGLGLHVAVSLFCNDK 260

Qy 121 NNNVRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIOGLGTIGQVSLRALRLRFYNE 180
Db 261 NNNVRALIRKDSGFVKASNLAKFELLEDLEFFYHFIOGLGTIGQVNLRLRFEYND 320

Qy 181 KQFFQFNYCSTLSKKFVFDNIM 204
Db 321 KQFFHFNYCNSLKKFVFDNIM 344

RESULT 3
Q8R695_FUSNN
ID Q8R695_FUSNN PRELIMINARY; PRT; 365 AA.
AC Q8R695;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-).
GN OrderedLocusNames=FN0560;
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
```

```
RX DOI=10.1128/JB.184.7.2005-2018.2002;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyripides N.C., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE009951; AAL94756.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0004109; F: coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F: iron ion binding; IEA.
DR GO; GO:0006779; P: porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN C.
DR InterPro; IPR004559; HemN rel.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF06969; HemN C; 1.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome; Oxidoreductase.
SQ SEQUENCE 365 AA; 43029 MW; 41EEF343A04309D6 CRC64;

Query Match 19.3%; Score 210; DB 2; Length 365;
Best Local Similarity 32.0%; Pred. No. 5.1e-10;
Matches 56; Conservative 32; Mismatches 61; Indels 26; Gaps 5;

Qy 1 RKFPFD-LNIDMTVMPLQKSHLRDLKELLSYMPHEICFSDPICBEGFVLRFDF--- 55
Db 147 RCGFDNISLDINFSLPYQTLSMLQNDLEKVLSPNPHISYLSIWEEGTKFFRDLKSGK 206

Qy 56 -----DNSIDSEKLWFCALCELSNGYINVTITPALKGHESRNKLNWELKPHGLGLY 111
Db 207 LKETDNDLEA-SMYEYIIIEFLKSDYIHEISNFKSDFESRNSTYWNKKYLGVLGA 265

Qy 112 VSLI-----FCNDKNNVRALIRK---TGSPVKANNHLVTFELLEDL 150
Db 266 AGYLNVRVYKNFNLKDYNNLDRNLIPDEKEILTEEDIEQRYLVGFRLINKI 320

RESULT 4
Q8R71_THETN
ID Q8R71_THETN PRELIMINARY; PRT; 375 AA.
AC Q8R71;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Coproporphyrinogen III oxidase and related FeS oxidoreductases.
GN Name=HemN; OrderedLocusNames=TT0952;
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336; DOI=10.1101/gr.219302;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of the T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013060; AAM24208.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0004109; F: coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F: iron ion binding; IEA.
DR GO; GO:0006779; P: magnesium ion binding; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006779; P: porphyrin biosynthesis; IEA.
DR InterPro; IPR002453; Beta_tubulin.
```


QY	TIGRFAMS; TIGR00539; hemN_rel; 1.	QY	TIGRFAMS; TIGR00539; hemN_rel; 1.
KW	Complete proteome.	KW	Complete proteome.
SEQ	SEQUENCE 374 AA; 43370 MW; B086EDD0D1491C84 CRC64;	SEQ	SEQUENCE 374 AA; 43370 MW; B086EDD0D1491C84 CRC64;
	Query Match 18.8%; Score 205; DB 2; Length 374;		Query Match 18.8%; Score 205; DB 2; Length 374;
	Best Local Similarity 32.1%; Pred. No. 1.4e-09;		Best Local Similarity 32.1%; Pred. No. 1.4e-09;
	Matches 62; Conservative 32; Mismatches 77; Indels 22; Gaps 8;		Matches 62; Conservative 32; Mismatches 77; Indels 22; Gaps 8;
QY	6 DLNIDMTVMNPLOKXGHLKRDLSYMEPHICFSDFICEE-----EGFVLRDFDN 57	QY	6 DLNIDMTVMNPLOKXGHLKRDLSYMEPHICFSDFICEE-----EGFVLRDFDN 57
DB	151 NINVDLMFGIPDQTLDFKESLEFITKLPKPHISSYSLIVBEGTGYFKMNEGGLK-LPN 209	DB	151 NINVDLMFGIPDQTLDFKESLEFITKLPKPHISSYSLIVBEGTGYFKMNEGGLK-LPN 209
QY	58 SIDSEKLMFALCCLNSNGHYINFTNFKLKGHSRINKLNWELKPHLGGLYAVSLFC 117	QY	58 SIDSEKLMFALCCLNSNGHYINFTNFKLKGHSRINKLNWELKPHLGGLYAVSLFC 117
DB	210 EDEBDMYSPARTFLEBKGGNQYEISNFAVKDKCEHNLNLYWELDNVYIGCCASAH--YF 267	DB	210 EDEBDMYSPARTFLEBKGGNQYEISNFAVKDKCEHNLNLYWELDNVYIGCCASAH--YF 267
QY	118 ND---KN-NVNRALIRK--TGSFVKANNHLVTFLELLEDFEFVVFHFGQLGTIGVSLRA 171	QY	118 ND---KN-NVNRALIRK--TGSFVKANNHLVTFLELLEDFEFVVFHFGQLGTIGVSLRA 171
DB	268 NGVRYRNNNNKKYIIQISKGNVVEENHRNL--LKEDMEEPMF---LGLRKTIRGVSIIE 322	DB	268 NGVRYRNNNNKKYIIQISKGNVVEENHRNL--LKEDMEEPMF---LGLRKTIRGVSIIE 322
QY	172 LRLRFBYNEKQFF 184	QY	172 LRLRFBYNEKQFF 184
DB	323 FKLKFNKDIOEY 335	DB	323 FKLKFNKDIOEY 335
	RESULT 6		RESULT 6
	Q892Q7 CLOTE		Q892Q7 CLOTE
ID	Q892Q7 CLOTE PRELIMINARY; PRT; 382 AA.	ID	Q892Q7 CLOTE PRELIMINARY; PRT; 382 AA.
AC	Q892Q7	AC	Q892Q7
DT	01-JUN-2003 (TREMBlrel. 24, Created)	DT	01-JUN-2003 (TREMBlrel. 24, Created)
DT	01-MAR-2004 (TREMBlrel. 24, Last sequence update)	DT	01-MAR-2004 (TREMBlrel. 24, Last sequence update)
DT	01-JUN-2004 (TREMBlrel. 26, Last annotation update)	DT	01-JUN-2004 (TREMBlrel. 26, Last annotation update)
DE	Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-)	DE	Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-)
GN	OrderedLocusNames=CTC02034;	GN	OrderedLocusNames=CTC02034;
OS	Clostridium tetani	OS	Clostridium tetani
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.	OC	Clostridium.
NCBI	NCBI_TaxID=1513;	NCBI	NCBI_TaxID=1513;
RN	[1]	RN	[1]
RP	NUCLEOTIDE SEQUENCE.	RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=Massachusetts / E98;	RC	STRAIN=Massachusetts / E98;
RX	MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;	RX	MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335853100;
RA	Brueggemann H., Baeumer S., Fricke W.F., Wierzer A., Liesegang H.,	RA	Brueggemann H., Baeumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA	Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,	RA	Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA	Gottschalk G.;	RA	Gottschalk G.;
RT	"The genome sequence of Clostridium tetani, the causative agent of	RT	"The genome sequence of Clostridium tetani, the causative agent of
RT	tetanus disease.";	RT	tetanus disease.";
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).	RL	Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).
RL	EMBL; AB015943; AAO36537.1; -; Genomic_DNA.	RL	EMBL; AB015943; AAO36537.1; -; Genomic_DNA.
DR	GO; GO:0005737; C:cytoplasm; IEA.	DR	GO; GO:0005737; C:cytoplasm; IEA.
DR	GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.	DR	GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR	GO; GO:0005506; F:iron ion binding; IEA.	DR	GO; GO:0005506; F:iron ion binding; IEA.
DR	GO; GO:0000287; F:magnesium ion binding; IEA.	DR	GO; GO:0000287; F:magnesium ion binding; IEA.
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.	DR	GO; GO:0016491; F:oxidoreductase activity; IEA.
DR	GO; GO:0006779; P:porphyrin biosynthesis; IEA.	DR	GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR	InterPro; IPR006638; EIP3/MiA/B/NifB.	DR	InterPro; IPR006638; EIP3/MiA/B/NifB.
DR	InterPro; IPR010723; HemN C.	DR	InterPro; IPR010723; HemN C.
DR	InterPro; IPR004559; HemN_rel.	DR	InterPro; IPR004559; HemN_rel.
DR	InterPro; IPR007197; Radical_SAM.	DR	InterPro; IPR007197; Radical_SAM.
DR	Pfam; PF06969; HemN C; 1.	DR	Pfam; PF06969; HemN C; 1.
DR	Pfam; PF04055; Radical_SAM; 1.	DR	Pfam; PF04055; Radical_SAM; 1.
DR	SMART; SM00729; EIP3; 1.	DR	SMART; SM00729; EIP3; 1.
DR	TIGRFAMS; TIGR00539; hemN_rel; 1.	DR	TIGRFAMS; TIGR00539; hemN_rel; 1.
KW	Complete proteome; Oxidoreductase.	KW	Complete proteome; Oxidoreductase.
SEQ	SEQUENCE 382 AA; 45516 MW; 52FA3CD797A6FFAA CRC64;	SEQ	SEQUENCE 382 AA; 45516 MW; 52FA3CD797A6FFAA CRC64;
	Query Match 17.7%; Score 192.5; DB 2; Length 382;		Query Match 17.7%; Score 192.5; DB 2; Length 382;
	Best Local Similarity 27.7%; Pred. No. 1.8e-08;		Best Local Similarity 27.7%; Pred. No. 1.8e-08;
	Matches 59; Conservative 42; Mismatches 83; Indels 29; Gaps 8;		

```

Qy 51 VLRFDFNSIDSEKLMFCALECLSGYINYEITNPFALKGHESRHNKLNWELKPHLGLGLY 110
Db 212 L--NLPNEDIERMYRVAIFLENGHGHQYEISNFSKREHCINLIYWDLSEYIGGCLA 269

Qy 111 AVSLL-----FCNDKNNVNRALIRKTSFVKANNHL-VTFELLEDLEFFVYHFYIQLGTTIQ 165
Db 270 AHSPLKGYRYSNVHNIEDIYKILINENKNIKINTYKNTKDTMEEFM-----MGLRKIK 323

Qy 166 GVSRLRALRPEYNEKQFFQFYNVCSLTKKFFV 198
Db 324 GINTEEFYKGFPHKNIEYV-----GDIIRKYYI 350

RESULT 7
Q7P440_FUSNV
ID Q7P440_FUSNV PRELIMINARY; PRT; 165 AA.
AC Q7P440;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Coproporphyrinogen oxidase, anaerobic (EC 1.-.-).
GN Name=FNV0357;
OS Fusobacterium nucleatum subsp. vincentii ATCC 49256.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=209882;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 49256;
RA Karpatrik V., Ivanova N., Anderson I., Reznik G., Bhattacharyya A.,
RA Gardner W L., Mikhailova N., Larsen N., D'Souza M., Walunas T.,
RA Habelkorn R., Overbeek R., Kyrpides N.;
RA Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABF01000140; EAA23389.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR010723; HemN_C.
DR Pfam; PF06969; HemN_C; 1.
DR Oxidoreductase.
KW OXIDOREDUCTASE.
SQ SEQUENCE 165 AA; 19650 MW; C4532CBBC40DAE01 CRC64;

Query Match 17.5%; Score 190.5; DB 2; Length 165;
Best Local Similarity 31.1%; Pred. No. 1.1e-08;
Matches 50; Conservative 30; Mismatches 56; Indels 25; Gaps 4;

Qy 14 NMPLQKSHLRDLKELLSYMPHEHICFSDPICEBEGFVLRFD-----DNSIDSEKLM 65
Db 3 SLPNQTWAMLQNDLEKLVNLPNPHISYLSIWEETKFFDLKSGKLKTDNDLEA-SMY 61

Qy 66 FCALECLSGYINYEITNPFALKGHESRHNKLNWELKPHLGLGLYAVSLL----- 115
Db 62 EYIEFLSKDYTHYEISNFSKKGPESSHNSIYENKNYLVGLSAGYLDNVRYKNFFN 121

Qy 116 ---FCNDKNNVNRALIRK---TGSPVKANNHLVTFELLEDL 150
Db 122 LKDYNNDKKNLIPIDEXEILTQEDIEQYRYLVGFRLLNKI 162

RESULT 8
Q8XIS8_CLOPE
ID Q8XIS8_CLOPE PRELIMINARY; PRT; 377 AA.
AC Q8XIS8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable coproporphyrinogen III oxidase.
GN Name=hemN; OrderedLocusNames=CPE2036;

OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=13 / Type A;
RC MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohehima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RA "Complete genome sequence of Clostridium perfringens, an anaerobic
RA flesh-eater.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
RL EMBL; BA000016; BAB81742.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; ELP3/MiaB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR04559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; ELP3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
DR Complete proteome.
KW COMPLETE PROTEOME.
SQ SEQUENCE 377 AA; 43713 MW; 0BCEFDAD409B34F78 CRC64;

Query Match 17.3%; Score 189; DB 2; Length 377;
Best Local Similarity 26.5%; Pred. No. 3.7e-08;
Matches 49; Conservative 37; Mismatches 83; Indels 16; Gaps 4;

Qy 6 DLNIDMTVMNPLQKSHLRDLKELLSYMPHEHICFSDPICEBEGFVLRFD-----S 58
Db 154 NINVDLMFGLPNQRLNWEETLREIISLEPAHISAYSLSLIEEGTAFYNLYENDKCLKPTE 213

Qy 59 IDSEKLMFCALECLSGYINYEITNPFALKGHESRHNKLNWELKPHLGLGLYAVSLLFCN 118
Db 214 EERKMVHLAKKILEENGFNQYEISNYSYKGEKGRHNLAYNMNDNWIWGVSAAS--YIN 271

Qy 119 DKN-----NNVRALIRKTSFVKANNHLVTFELLEDLEFFVYHFYIQLGTTQGVSLRALRL 174
Db 272 GKRIKNISSVEKYINSINEKREAVEEIIINNSKNDNMEEFME--MGLRKINGIDENEFKN 328

Qy 175 RFEYN 179
Db 329 RFSMN 333

RESULT 9
Q8YMH7_ANASP
ID Q8YMH7_ANASP PRELIMINARY; PRT; 396 AA.
AC Q8YMH7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Coproporphyrinogen III oxidase.
GN OrderedLocusNames=all4956;
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yanada M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RA cyanobacterium Anabaena sp. strain PCC 7120.";
RT

```

DR Pfam; PF04055; Radical_SAM; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
DR Complete proteome.
SQ SEQUENCE 408 AA; 45119 MW; 936C6CD8C5992C39 CRC64;

Query Match 16.1%; Score 175; DB 2; Length 408;
Best Local Similarity 27.9%; Pred. No. 6.8e-07;
Matches 53; Conservative 27; Mismatches 76; Indels 34; Gaps 5;

QY 6 DLNIDMTVNNPLQCKSHLKRDLKELLSYMPHEHICFSDFICEEFGFVLRF-----DNS 58
DB 187 NFNLDLIFGLPHQSLADWEFSLEVLRLFPETHLSLYDLILEETPFGRPLGGQAPLPGD 246
QY 59 IDEKULWFCALCELENGYINYEITPALKGHESRNKLNWELKPHLGLGLYAVSLFCN 118
DB 247 DDTVAMYLACERLQCGYRQEYIANFARPGFCRHRVVDNRPPYVGIG-----296
QY 119 DKNNVRALIRKTSFVKANNHLLVT--PELLEDEF-----FVHFIOGLGTIOG 166
DB 297 ---NGATGVYR--GARVERPRLLDHYFAWVEGGEFFAGAPVMPPEELADTLILGLRLEG 351
QY 167 VLSRALRLRF 176
DB 352 VEVNALVQRF 361

RESULT 11
Q73MP2_TREDE
ID Q73MP2_TREDE PRELIMINARY; PRT; 391 AA.
AC Q73MP2;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase, putative.
GN OrderedLocusNames=TDE1466;
OS Treponema denticola.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=158;
RN [1]
NP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35405 / DSM 14222;
RX PubMed=15064399; DOI=10.1073/pnas.0307639101;
RA Seshadri R., Myers G.S.A., Tettelin H., Eisen J.A., Heidelberg J.F.,
RA Dodson R.J., Davidsen T.M., DeBoy R.T., Fouts D.E., Haft D.H.,
RA Selengut J., Ren Q., Brinkac L.M., Madupu R., Kolonay J.F.,
RA Durkin S.A., Daugherty S.C., Shetty J., Shvartsbeyn A.,
RA Gebrgeorgis E., Geer K., Tsegaye G., Malek J.A., Ayodeji B.,
RA Shatsman S., McLeod M.P., Smajs D., Howell J.K., Pal S., Amin A.,
RA Vashisth P., Neilell T.Z., Xiang Q., Sodergren E., Baca E.,
RA Weinstein G.M., Norris S.J., Fraser C.M., Paulsen I.T.;
RT "Comparison of the genome of the oral pathogen Treponema denticola
with other spirochete genomes".
RL Proc. Natl. Acad. Sci. U.S.A. 101:5646-5651(2004).
DR EMBL; AE017251; AAS11983.1; -; Genomic_DNA.
DR TIGR; TDE1466; -
DR GO; GO:0005737; C:cyttoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
DR Complete proteome.
SQ SEQUENCE 391 AA; 45034 MW; 8D121969939FF1CB CRC64;

Query Match 15.8%; Score 172.5; DB 2; Length 391;


```
Db 320 RFKKKFDQSIESVPGQTIN--NLKEKKLIVEKN 350
RESULT 14
Y498 BUCBP
ID Y498 BUCBP STANDARD; PRT; 374 AA.
AC Q89A47;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical protein bhp498.
GN OrderedLocusNames=bbp498;
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22426901; PubMed=12522265; DOI=10.1073/pnas.0235981100;
RA van Ham R.C.H.J., Kamerbeek J., Palacios C., Raueell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003).
CC -1- SIMILARITY: Belongs to the anaerobic coproporphyrinogen III
CC oxidase family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AB014017; AA027203.1; -; Genomic_DNA.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome; Hypothetical protein; Oxidoreductase;
KW Porphyrin biosynthesis.
SQ SEQUENCE 374 AA; 44752 MW; 9D93C81AD2565053 CRC64;
Query Match 14.4%; Score 157.5; DB 1; Length 374;
Best Local Similarity 22.5%; Pred. No. 2.1e-05;
Matches 45; Conservative 35; Mismatches 73; Indels 47; Gaps 6;
Qy 6 DLNIDMTVMPLQKSHLRDLKELLSYMPHICFSDPICEEGFVLDRDNDSDSKLW 65
Db 156 NINLDLYGLPKQSLQEQALLDLKTAISLKPNIHSWCFYEIKNNNNYKNLSKSCNLIW 215
Qy 66 FCALE---CLESNGYINVEITNPAKLGHSRHNKLNWELKPHGLGLYAVSLFLCNDKN 122
Db 216 KIFLQGEKLLKSKYKKEYSYKNTYQCLNHNLYWKFDYLGIG-----CNAHGK 267
Qy 123 NVRALIRKTSFVKA-----NNHLVTFELLEDLEFFVHFHFIQGLGIQ 166
Db 268 -----ITQNGKIITKNNKLNKFPWNGKYTYKNHISKNL-SLEFF----- 309
Qy 167 VSLRALRLRFEYNEKQFFQF 186
Db 310 --NNRLRLNTPYRKDFKY 327
RESULT 15
QSLZC1 STRT1 PRELIMINARY; PRT; 378 AA.
AC QSLZC1;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Coproporphyrinogen III oxidase.
```

```
GN Name=hemN; OrderedLocusNames=stl1236;
OS Streptococcus thermophilus (strain CNRZ 1066).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=299768;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15543133; DOI=10.1038/nbt1034;
RA Bolotin A., Quinkis B., Renault P., Sorokin A., Ehrlich S.D.,
RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,
RA Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcour J.,
RA Goffeau A., Hols P.;
RT "Complete sequence and comparative genome analysis of the dairy
RT bacterium Streptococcus thermophilus.";
RL Nat. Biotechnol. 22:1554-1558(2004).
DR EMBL; CP000024; AAV62781.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; ELP3/MiAB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; ELP3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome.
SQ SEQUENCE 378 AA; 43487 MW; 450DD78F16B0D966 CRC64;
Query Match 14.1%; Score 154; DB 2; Length 378;
Best Local Similarity 26.9%; Pred. No. 4.3e-05;
Matches 59; Conservative 33; Mismatches 83; Indels 42; Gaps 10;
Qy 6 DLNIDMTVMPLQKSHLRDLKELLSYMPHICFSDPICEEGFVLDRD-----DN 57
Db 153 NISIDLIYALPGQTMQVKENVKALELDIPHLSLYSLLLEHHTVFMNKNRGGKLNLPTE 212
Qy 58 SIDSEKLNWFCALCLESNGYINVEITNPAKLGHSRHNKLNWELKPHGLGLYAVSLFLC 117
Db 213 DLEAEMFDYTIQI--LEKNGFHEYEISNFTKPGMESRHNLMYNNDEYIGVGAGAGTV-- 269
Qy 118 NDKNNNVRLIRKTG---SFVKA-----NNHLVTFELLEDLEFFVHFHFIQGLGTI 164
Db 270 ----NGVR--YRNRGPIQHYLKAIADGDHARLNNEEHLTKVEMWEE--EFPL-----GLRKK 317
Qy 165 QGVSLRALRLRFEYNEKQFFQFINYCGTSLSKKFVFD 200
Db 318 SGVSIK----RFE--EKFGLSFSDTYGDIVKKLQED 347
Search completed: January 24, 2006, 19:56:46
Job time : 45.8352 secs
```

This Page Blank (uspio)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	145	13.3	155	2	US-09-830-230A-522	Sequence 522, App
2	145	13.3	175	2	US-09-830-230A-521	Sequence 521, App
3	134	12.3	376	2	US-09-583-110-3497	Sequence 3497, Ap
4	134	12.3	409	2	US-09-107-433-4312	Sequence 4312, Ap
5	112	10.3	391	2	US-09-134-001C-3690	Sequence 3690, Ap
6	111	10.2	393	2	US-09-107-532A-6801	Sequence 6801, Ap
7	105.5	9.7	401	2	US-09-902-540-14907	Sequence 14907, A
8	105	9.6	424	2	US-09-252-991A-20695	Sequence 20695, A
9	104.5	9.6	397	2	US-09-134-000C-5885	Sequence 5885, Ap
10	102.5	9.4	394	2	US-09-198-452A-396	Sequence 396, App
11	102.5	9.4	402	2	US-09-438-185A-382	Sequence 382, App
12	98	9.0	387	2	US-09-489-039A-11948	Sequence 11948, A
13	96.5	8.9	382	2	US-09-543-681A-5208	Sequence 5208, Ap
14	93	8.5	5037	2	US-09-242-783-4	Sequence 4, Appli
15	88.5	8.1	287	2	US-09-830-230A-122	Sequence 122, App
16	88.5	8.1	309	2	US-09-830-230A-121	Sequence 121, App
17	88.5	8.1	387	2	US-09-248-796A-22214	Sequence 22214, A
18	88	8.1	5032	2	US-09-538-092-979	Sequence 979, App
19	86	7.9	389	2	US-09-328-352-8088	Sequence 8088, Ap
20	85	7.8	460	2	US-09-198-452A-959	Sequence 959, App
21	85	7.8	460	2	US-09-438-185A-891	Sequence 891, App
22	82	7.5	694	2	US-09-538-092-1351	Sequence 1351, Ap
23	82	7.5	698	2	US-09-949-016-10215	Sequence 10215, A
24	81.5	7.5	1255	2	US-09-248-796A-14158	Sequence 14158, A
25	81	7.4	254	2	US-09-134-000C-6525	Sequence 6525, Ap
26	81	7.4	268	2	US-09-543-681A-7944	Sequence 7944, Ap
27	81	7.4	899	2	US-09-248-796A-15467	Sequence 15467, A

; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 521
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-521

Query Match 13.3%; Score 145; DB 2; Length 175;
Best Local Similarity 90.0%; Pred. No. 2.3e-08;
Matches 27; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RKFPFDLNDMTVNMPLQKSHLRDLKEL 30
Db 141 RKFPFDLNDMTVNMPLQKSHLRDLQRI 170

RESULT 3
US-09-583-110-3497
; Sequence 3497, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 3497
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-3497

Query Match 12.3%; Score 134; DB 2; Length 376;
Best Local Similarity 26.4%; Pred. No. 1.2e-06;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVNMPLQKSHLRDLKELSYNPEHICFSDFCICEBEGFVLRDF----- 55
Db 151 FDNISIDLIALPGQTMQVKENVAKAIGLDIPHMSLYSLILENHTVPMNMRGKPLP 210

Qy 56 DNSIDSEKLFWCALECLSNYINYEITNPAKKGHESRHNKLWELKPHLGLGLYAVSLL 115
Db 211 KEELEAEMFYITIAE-LERAGFEHYEISNFKPGFESRHNLMYWDNAEYYIGAGAGYV 269

Qy 116 FCNDKNNVNR---ALIRKTSFVKANNHLVTFFLEDFVYHFIOGLGTIOGVSLRA 171
Db 270 -----NGVRYKHGPIRHYLSAVEEGNACITEDHLSQKEQMEEMFLGLRKKSGVSM-- 321

Qy 172 LRLRFEYNEKQFF 184
Db 322 --ARFEKFGQSF 332

RESULT 4
US-09-107-433-4312
; Sequence 4312, Application US/09107433
; Patent No. 6800744

; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4312:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...409
; SEQUENCE DESCRIPTION: SEQ ID NO: 4312:
US-09-107-433-4312

Query Match 12.3%; Score 134; DB 2; Length 409;
Best Local Similarity 26.4%; Pred. No. 1.4e-06;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVNMPLQKSHLRDLKELSYNPEHICFSDFCICEBEGFVLRDF----- 55
Db 184 FDNISIDLIALPGQTMQVKENVAKAIGLDIPHMSLYSLILENHTVPMNMRGKPLP 243

Qy 56 DNSIDSEKLFWCALECLSNYINYEITNPAKKGHESRHNKLWELKPHLGLGLYAVSLL 115
Db 244 KEELEAEMFYITIAE-LERAGFEHYEISNFKPGFESRHNLMYWDNAEYYIGAGAGYV 302

Qy 116 FCNDKNNVNR---ALIRKTSFVKANNHLVTFFLEDFVYHFIOGLGTIOGVSLRA 171
Db 303 -----NGVRYKHGPIRHYLSAVEEGNACITEDHLSQKEQMEEMFLGLRKKSGVSM-- 354

Qy 172 LRLRFEYNEKQFF 184
Db 355 --ARFEKFGQSF 365

RESULT 5
US-09-134-001C-3690
; Sequence 3690, Application US/09134001C
; Patent No. 6380370


```

; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6801:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (B) LOCATION 1...393
; SEQUENCE DESCRIPTION: SEQ ID NO: 6801:
US-09-107-532A-6801

Query Match 10.2%; Score 111; DB 2; Length 393;
Best Local Similarity 20.7%; Pred. No. 0.00054;
Matches 42; Conservative 37; Mismatches 88; Indels 36; Gaps 5;

QY 6 DLNIDVTNNMPLQKSHLKRDLKELLSYMEPHICFSDFICEEGFVLK-----DFDNS 58
DB 168 NVSIDLIYAUPGQGLEGYLTLEKALALDLPKHSLSVLSLIENKTMFMNWRQGRLELPDQ 227
QY 59 IDSEKLFWCALECLSENGYINYYITTFALPKGHESRRHKNLWELKPHLGLGLYAVSLFLFCN 118
DB 228 ETETRMFEETIEAMKXGRHRYEISNFGLSGHESQHNLMYWNNDHYFGGAGASGYL--- 284
QY 119 DKNNVRVALIRKTGSFVKANNHLVTFELLEDLBFFVYHFHIOGLGTIOGVSLRALRL-RPE 177
DB 285 ---GDIR--YRNKGP-----IQHYLRPLRANRPPVLEBEKLRXN 319
QY 178 YNEKQFFQFINYCSTLSKKFVFD 200
DB 320 QIEEMFLGURKKTGITKXHFYD 342

RESULT 7
US-09-902-540-14907
; Sequence 14907, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 14907
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-14907

Query Match 9.7%; Score 105.5; DB 2; Length 401;
Best Local Similarity 23.2%; Pred. No. 0.0023;
Matches 47; Conservative 28; Mismatches 83; Indels 45; Gaps 7;

QY 3 FPDFLIDMTVMNPLQKSHLKRDLKELLSYMEPHICFSDPICEEGFVLK----- 54
DB 170 FVPV-VAMDFYGVHGQTVQAVEADARRAVALAPFHL--STYALTVERDVLAEIDTFLSKRL 226
QY 55 -----FDNSIDSEKLFWCALECLSENGYINYYITTFALPKGHESRRHKNLWELKPHL 105
DB 227 KRGELELPDDTVDVDMARV---VREYVYAGLRHYEVSNNHARFGFSRRHNLWYTGGEYL 283
QY 106 GIGLYAVSLILFCNDKKN--NVRALIRKTGSFVKANNHLVTFELLEDLBFFVYHFHIOGLGTIOGVSLRALRL-RPE 177

```

```
Db      284  ALGVGATGMLSPQPHRYNLS-----AERYVEAEAGRLPEEGREPLGPPELF 333
Qy      154  VYHFIOGLGTIQGVSLRALRLRF 176
Db      334  AERLAWGLRLVSGVDWEAVCERY 356

RESULT 8
US-09-252-991A-20695
; Sequence 20695, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20695
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20695

Query Match          9.6%; Score 105; DB 2; Length 424;
Best Local Similarity 26.0%; Pred. No. 0.0029;
Matches 32; Conservative 18; Mismatches 69; Indels 4; Gaps 2;

Qy      1  RKFPFD-LNIDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFICEEGFVLRDFDMSI 59
Db      198  RQAGFDNFDLMHGLPQSLDDALDLRQALAPATHLSVQLTMEPTVWQPPPELP 257
Qy      60  DSEKLN----FCALECLSENGYINYEITNFALKGHESRNKLNWELKPHLGLYAVSLLF 116
Db      258  EDDLWDIQEAGQALLDHEGYQVETSAYARDGLHARHNLNYSFGDFLGIGAGAHAKLS 317
Qy      117  CND 119
Db      318  APD 320

RESULT 9
US-09-134-000C-5885
; Sequence 5885, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5885
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5885

Query Match          9.6%; Score 104.5; DB 2; Length 397;
Best Local Similarity 26.0%; Pred. No. 0.003;
Matches 33; Conservative 14; Mismatches 53; Indels 27; Gaps 2;

Qy      6  DLNIDMTVMPLQKSHLKRDLKELLSYMPHEICFSDFICE-----EE 48
```

```
Db      170  NVSIDLIYALPGQTLSEFRDTLTRALALDLPFHYSLYSLILENKTMTFMNWMVQRQLQLPPE 229
Qy      49  GFVLRDFDMSIDSEKLFWFCALCELESGYINYEITNFALKGHESRNKLNWELKPHLGLG 108
Db      230  EIEAQMFDETI-----EAMEKKGRRHQYEVSNFALTGKESQHNLAYWNNDHYHYGFG 279
Qy      109  LYAVSLL 115
Db      280  AGASGYL 286

RESULT 10
US-09-198-452A-396
; Sequence 396, Application US/09198452A
; Patent No. 6552994
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 396
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-396

Query Match          9.4%; Score 102.5; DB 2; Length 394;
Best Local Similarity 21.9%; Pred. No. 0.005;
Matches 49; Conservative 40; Mismatches 96; Indels 39; Gaps 8;

Qy      6  DLNIDMTVMPLQKSHLKRDLKELLSYMPHEICFSDF-----ICEE 47
Db      175  NLSIDLIYGLPTQSLSEIFLSDLHQALTPTTHISLYNLTDIDPHTSFYKRRKILVPTTIAQE 234
Qy      48  EGFVLRDFDMSIDSEKLFWFCALCELESGYINYEITNFALKGHESRNKLNWELKPHLGL 107
Db      235  E--ILAEH--SLIAENL-----LLSQGFQRYELASYAKPDYPAKHNLYYTWDRPFJGL 283
Qy      108  GLYAVSLLFCNDKNN-----VRALIRKTGSPVKANNHLVTFELLEDLEFFVVFHTQ 160
Db      284  GVSASQYLH-GERSKNYSHISHYLRA-VRKNLPQETSEILPKKERIKEALALRLLEG 341
Qy      161  LGTIQGVSLRALRLRFEYNEKQFFQFINYCSTLSKK-FVFDNI 203
Db      342  ADLAEFPSTLISMLTQDVKLQNLPSVHGQCCLALNRQGLRPHDTI 385

RESULT 11
US-09-438-185A-382
; Sequence 382, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 382
```

```
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5208
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5208

Query Match      8.9%; Score 96.5; DB 2; Length 382;
Best Local Similarity 25.0%; Pred. No. 0.023;
Matches 26; Conservative 17; Mismatches 58; Indels 3; Gaps 1;

Qy      8 NIDWTNMPLOKXSHLKDULKELLSSYMPEHICFSDFTCEBEGFVLRFDFDNSIDSEKLWFC 67
Db      167 NLDLMEGLFPQSLSQALSLSLQAIALSPHLHWQLTIENTQFSRPPKLPDDMLNDI 226

Qy      68 ALE---CLENGYINYEITNFALKGHESRHKNLNWLKPHLG LG 108
Db      227 FSEGDKLLTAAGYQVETSAYCKPGFCQCQHNLNWRFGDYLGIG 270

RESULT 14
US-09-424-783-4
Sequence 4, Application US/09424783
Patent No. 6780608
GENERAL INFORMATION:
APPLICANT: Hakamata, Yasuhiro
APPLICANT: Nishimura, Seiichiro
APPLICANT: Barsoumian, Edward Leon
TITLE OF INVENTION: Human Type 3 Ryanodine Receptor Protein
TITLE OF INVENTION: and DNA Molecules Coding Therefor
FILE REFERENCE: 0652.2000000
CURRENT APPLICATION NUMBER: US/09/424,783
CURRENT FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/EP98/02926
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: DE 197 22 317.6
PRIOR FILING DATE: 1997-05-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent in version 3.1
SEQ ID NO 4
LENGTH: 5037
TYPE: PRT
ORGANISM: Oryctolagus cuniculus
US-09-424-783-4

Query Match      8.5%; Score 93; DB 2; Length 5037;
Best Local Similarity 24.0%; Pred. No. 2.3;
Matches 49; Conservative 36; Mismatches 75; Indels 44; Gaps 11;

Qy      13 VNPMPLQKKSHLKRD LK--ELLSYNPEHI-----CFSDFTCEBEGFV-LRDFNSIDSEKLWF 66
Db      4045 VDMLVESSNVNEMILKFDMFLKLDKIVGSEAFQDYVTDPRGLISKDKDFQKAMDSQKQT 4104

Qy      67 -----CALECLES--NCYINVEITNFALKGHESRHKNLNWLKPHLG LGLYAVSLLFCND 119
Db      4105 GPETQFFLLSCSEADENEMINF--EFANRPQE-----PARDIGF----- 4141

Qy      120 KNNVRALIRKTGSFVKANNHLVTFELEDLEFFVHFIOGLGTIQ--GVSLRALRLRFE 177
Db      4142 ---NVAVLLTNLSEHVHPDRLRNF--LELAESILEYFRPYLGRIEIMGASRRIERIYFE 4196

Qy      178 YNEKQFFQF--INYCSTLSKXFVD 200
Db      4197 ISETNRAQWEMPQVKESKRQFI D 4220

RESULT 15
US-09-830-230A-122
```

```
; Sequence 122, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-122

Query Match      8.1%; Score 88.5; DB 2; Length 287;
Best Local Similarity 22.7%; Pred. No. 0.13;
Matches 45; Conservative 25; Mismatches 67; Indels 61; Gaps 7;

Qy      3 FPPDLNIDMTVMPLQKKSHLRDLKELLSYMPHEHICFSDPICEEGFVLRLDFDINSIDSE 62
Db      14 FLFKIN-----KSLGK---KELPIDQNTHCUSPEYDNLAKILIWDFKNELRKE 59

Qy      63 KLWFCALCLESNGYINYEITNFALKGHESHNNKLNWELKPHLGLGLYAVSLLFCNDKNN 122
Db      60 -----GFFTQQIKN-----DSSQYINARKN 79

Qy      123 NVRALIRKTGS---FVKANHILVTFELLEDL-EFFVYHFIOGLGTIOGVSLRALRLRPEY 178
Db      80 NISFSIKREGSKITFECFNNHLI---IIQDLFRETLNLEKITKEVETVSLRAKKLDYSI 136

Qy      179 NEKQFFQFINYCSTLSKK 196
Db      137 NYDKILSNINLNKRIKKE 154
```

Search completed: January 24, 2006, 20:02:41
Job time : 13.5762 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 35.8542 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-10
Perfect score: 1090
Sequence: 1 RKPPFDLNDMTVMPLQKK.....QFINYCSTLSKKKVFDDNIM 204

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.Main.*
1: /cgn2_5/ptodata/1/pubpaa/US07_PUBCOMB.pap.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1090	100.0	204	5	US-10-688-058-10
2	1090	100.0	377	5	US-10-688-058-12
3	151.5	13.9	385	5	US-10-470-048B-106
4	145	13.3	155	5	US-10-994-726-522
5	145	13.3	175	5	US-10-994-726-521
6	135	12.4	381	5	US-10-650-274-98
7	134	12.3	376	5	US-10-472-928-2860
8	134	12.3	409	5	US-10-617-320-4312
9	117	10.7	345	4	US-10-335-977-6717
10	117	10.7	352	4	US-10-335-977-6718
11	117	10.7	364	4	US-10-335-977-6719
12	104.5	9.6	493	4	US-10-425-115-261185
13	102.5	9.4	394	4	US-10-289-762-396
14	97	8.9	548	4	US-10-724-972A-7488
15	93	8.5	350	6	US-11-097-143-13632
16	91.5	8.4	605	4	US-10-282-122A-54636
17	91	8.3	1103	4	US-10-282-122A-76866
18	89.5	8.2	4473	4	US-10-437-963-201113
19	89.5	8.2	4737	4	US-10-437-963-201116
20	89	8.2	665	4	US-10-437-963-172583
21	88.5	8.1	287	5	US-10-994-726-122
22	88.5	8.1	309	5	US-10-994-726-121
23	88.5	8.1	701	4	US-10-282-122A-47140
24	88	8.1	1032	4	US-10-607-631-4
25	88	8.1	5032	4	US-10-408-765A-26
26	88	8.1	5038	5	US-10-723-860-1310
27	88	8.1	5038	5	US-10-756-149-5039

```

28      88      8.1  5081      4  US-10-276-774-1850      Sequence 1850, Ap
29      88      8.1  5081      5  US-10-450-763-35328      Sequence 35328, A
30      87      8.0  605      4  US-10-238-075-691      Sequence 691, App
31      85.5     7.8  463      6  US-11-097-143-14868      Sequence 14868, A
32      85      7.8  460      4  US-10-289-762-959       Sequence 959, App
33      84.5     7.8  204      4  US-10-335-977-7414      Sequence 7414, Ap
34      84.5     7.8  742      4  US-10-437-963-109290    Sequence 109290,
35      82.5     7.6  330      3  US-09-738-626-6014      Sequence 6014, Ap
36      82.5     7.6  410      4  US-10-156-761-13098     Sequence 13098, A
37      82.5     7.6  580      6  US-11-097-143-13371     Sequence 13371, A
38      82.5     7.6  739      4  US-10-437-963-155552    Sequence 155552,
39      82      7.5  483      4  US-10-437-963-125107    Sequence 125107,
40      82      7.5  694      3  US-09-842-758-75        Sequence 75, Appl
41      82      7.5  694      3  US-09-855-828-14        Sequence 14, Appl
42      82      7.5  694      4  US-10-345-680-26        Sequence 26, Appl
43      82      7.5  694      4  US-10-174-333-75        Sequence 75, Appl
44      82      7.5  694      5  US-10-978-282-11        Sequence 11, Appl
45      82      7.5  1067     4  US-10-437-963-121534    Sequence 121534,

```

ALIGNMENTS

RESULT 1

```

US-10-688-058-10
; Sequence 10, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 10
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-10

```

```

Query Match      100.0%; Score 1090; DB 5; Length 204;
Best Local Similarity 100.0%; Pred. No. 9.4e-106;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  RKPPFDLNDMTVMPLQKKSHLKRDLKELLSYMPHICFSDFCIEEGVFLRDFDMSID 60
      |||
Db      1  RKPPFDLNDMTVMPLQKKSHLKRDLKELLSYMPHICFSDFCIEEGVFLRDFDMSID 60
      |||

Qy      61  SEKLWFCALCELSNGYINYEITNFALKGHESHKLNKLKHLGLYAVSLFCNDK 120
      |||
Db      61  SEKLWFCALCELSNGYINYEITNFALKGHESHKLNKLKHLGLYAVSLFCNDK 120
      |||

Qy      121  NNNVRALIRKTSFVKANNHLVTFFELLEDFEYVYHFIQGLGTIQGVSLRALRLRFEYNE 180
      |||
Db      121  NNNVRALIRKTSFVKANNHLVTFFELLEDFEYVYHFIQGLGTIQGVSLRALRLRFEYNE 180
      |||

Qy      181  KQFFQFINYCSTLSKKKVFDDNIM 204
      |||
Db      181  KQFFQFINYCSTLSKKKVFDDNIM 204
      |||

```

RESULT 2

```

US-10-688-058-12
; Sequence 12, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.

```

APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCR0:0030US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-12

Query Match 100.0%; Score 1090; DB 5; Length 377;
Best Local Similarity 100.0%; Pred. No. 2.1e-105;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RKPPFDLNDMTVNMPLQKSHLKRDLKELLSYMPHEICFSDFTCEBEGFVLRDPFNSID 60
Db 141 RKPPFDLNDMTVNMPLQKSHLKRDLKELLSYMPHEICFSDFTCEBEGFVLRDPFNSID 200
Qy 61 SEKLWFCALCELESGNGYINVTNFALKGHESRHNKLWELKPHLGLGLYAVSLLFCNDK 120
Db 201 SEKLWFCALCELESGNGYINVTNFALKGHESRHNKLWELKPHLGLGLYAVSLLFCNDK 260
Qy 121 NNNVALIRKTGTFVKANNHLVTPELLEDFEYFVHFYHFIQGLGTTCQVSLRALRFRFYNE 180
Db 261 NNNVALIRKTGTFVKANNHLVTPELLEDFEYFVHFYHFIQGLGTTCQVSLRALRFRFYNE 320
Qy 181 KQFFQFNYCSTLKKFVDDNIM 204
Db 321 KQFFQFNYCSTLKKFVDDNIM 344

RESULT 3
US-10-470-048B-106
; Sequence 106, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; TITLE OF INVENTION: ANTIGENS TO A SPECIFIC PATHOGEN
; FILE REFERENCE: SONN:0350US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 106
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-106

Query Match 13.9%; Score 151.5; DB 5; Length 385;
Best Local Similarity 26.3%; Pred. No. 5.2e-07;
Matches 56; Conservative 44; Mismatches 84; Indels 29; Gaps 9;

Qy 7 LNIDMTVNMPLQKSHLKRDLKELLSYMPHEICFSDFTCEBEGFVLRDPFNSID 56
Db 161 ISLDLMLHLPKQTTEDFQSDLDLMDIQHISVGLILEPTQFYNNYRKGLLKLPNED 220
Qy 57 NSIDSEKLFWCALECSNGYINVTNFALKGHESRHNKLWELKPHLGLGLYAVSLLFCNDK 111
Db 221 LGADMYQL---LMSKIEQSPHQEISNFALDGHSEHNKVVWFNEEYFGCAGASGYVD 277
Qy 112 -VSLLFCDKNNVALIRKTGTFVKANNHLVTPELLEDFEYFVHFYHFIQGLGTTCQVSLR 170
Db 278 GVRVTNIPNVHYTKAINKSKAILVSNKPSLT-ERMEE-EMFL-----GLRLNEGVS 330
Qy 171 ALRLRFEYN-EKQFFQFNYCSTLKKFVDDNIM 202

Db 331 RFKKKFDQSIQSVFGQTIN--NLKKEKELIVEKN 361

RESULT 4
US-10-994-726-522
; Sequence 522, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 522
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-994-726-522

Query Match 13.3%; Score 145; DB 5; Length 155;
Best Local Similarity 90.0%; Pred. No. 7.5e-07;
Matches 27; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RKPPFDLNDMTVNMPLQKSHLKRDLKEL 30
Db 121 RKPPFDLNDMTVNMPLQKSHLKRDLQRI 150

RESULT 5
US-10-994-726-521
; Sequence 521, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 521
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-994-726-521

Query Match 13.3%; Score 145; DB 5; Length 175;

Best Local Similarity 90.0%; Pred. No. 8.8e-07;
Matches 27; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RKFPFDLNIDMTVMNPLQKKSHLRDLKEL 30
Db 141 RKFPFDLNIDMTVMNPLQKKSHLRDLQRI 170

RESULT 6
US-10-650-274-98
; Sequence 98, Application US/10650274
; Publication No. US20050202437A1
; GENERAL INFORMATION:
; APPLICANT: GLENN, MATTHEW
; APPLICANT: HAVUKKALA, ILKKA J
; APPLICANT: LUBBERS, MARK WILLIAM
; APPLICANT: DEKKER, JAMES
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES,
; TITLE OF INVENTION: MATERIALS INCORPORATING THEM, AND METHODS FOR USING
; TITLE OF INVENTION: THEM.
; FILE REFERENCE: 11000.1073
; CURRENT APPLICATION NUMBER: US/10/650.274
; CURRENT FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0
; SEQ ID NO 98
; LENGTH: 381
; TYPE: PRT
; ORGANISM: LACTOBACILLUS RHAMNOSUS
US-10-650-274-98

Query Match 12.4%; Score 135; DB 5; Length 381;
Best Local Similarity 26.2%; Pred. No. 2.8e-05;
Matches 53; Conservative 33; Mismatches 76; Indels 40; Gaps 8;

Qy 1 RKFPFD-LNIDMTVMNPLQKKSHLRDLKELLSYMPHEICFSDPFCIEEGFVLDRFNSI 59
Db 147 RKGVDNLSIDLIFRLPQSDRDFNSLQKALADLDPHYSTYSILERKTI----FYNLM 202

Qy 60 DSEKL-----WFCALCELESNGYINYEITNFALKGHESRHNKLNWELKPHLGLG 108
Db 203 RQRKLRLPTQDVEADMYQDAIDLEAGHGHQYELSNFAKTYQCRHNLLYWQNDKYFGF 262

Qy 109 LYAVSLFCNDKNNVRALIRKTSF-----VKANNHLVTFLELLEDDLEFPVYHFIOGLG 162
Db 263 AGAFGYL-GRDVRNYGPIKQYLAPLHADHLPVLAHLVPVSEQIEB-EMEL-----GLR 315

Qy 163 TIQGVSLRALRLRFEYNEKQFF 184
Db 316 TMAGV-----NEDRFY 326

RESULT 7
US-10-472-928-2860
; Sequence 2860, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472.928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 2860
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: oxygen-independent coproporphyrinogen III oxidase, putative

; OTHER INFORMATION: Cellular location: cytoplasm
; OTHER INFORMATION: Similar to strain R6 sequence 15903309 (O.E+01)
US-10-472-928-2860

Query Match 12.3%; Score 134; DB 5; Length 376;
Best Local Similarity 26.4%; Pred. No. 3.4e-05;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVMNPLQKKSHLRDLKELLSYMPHEICFSDPFCIEEGFVLDRDF----- 55
Db 151 FDNISIDLIALPQQTMEQVKENVAKAIGDIDPHMSLYSLILENHTVFMNMRGRGLPLP 210

Qy 56 DNSIDSEKLFWCALECLESNGYINYEITNFALKGHESRHNKLNWELKPHLGLGLYAVSL 115
Db 211 KEELEAEFYYIIAE-LERAGFHEYELSNFKGFSRHNMTWDNNAEYIGIGAGASYV 269

Qy 116 FCNDKNNVR-----ALIRKTSFVKANNHLVTFLELLEDDLEFPVYHFIOGLGTQGVSLRA 171
Db 270 -----NGVYKNGHPIRHYLSAVESGNACITEDHLSQKEQMEEMFLGLRKKSQVSM-- 321

Qy 172 LRLRFEYNEKQFF 184
Db 322 --ARFEKFGQSF 332

RESULT 8
US-10-617-320-4312
; Sequence 4312, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617.320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107.433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4312:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae

```
;
;
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...409
; SEQUENCE DESCRIPTION: SEQ ID NO: 4312:
US-10-617-320-4312

Query Match      12.3%; Score 134; DB 5; Length 409;
Best Local Similarity 26.4%; Pred. No. 3.8e-05;
Matches 51; Conservative 28; Mismatches 90; Indels 24; Gaps 6;

Qy 5 FD-LNIDMTVMPLQKSHLKRDLKELLSYMP-EBHICFSDFCICEEGFVLRDF-----55
Db 184 FDNISIDLIALPQTMQEVKENVAKAIGLDIPHMSLYSLILENHTVFMNRRRGKLP 243
Qy 56 DNSIDSEKLFCALECLSGYINYEITNFALKGHESRHNKLNWELKPHLGLGYAVSLL 115
Db 244 KEELEAEAFYIIAB-LERAGFEHYEISNFKSPGFESRHNLMYWDNABYYGIGAGASYV 302
Qy 116 FCNDKNNNRV---ALIRKGTGSFVKANNHLVTFPELLEDFVYHFTQGLGTIQGVSLRA 171
Db 303 -----NGVRYKNHGPRIHYLSAVEEGNACITEDHLSQKEQMEEMFLGLRKKSGVSM-- 354
Qy 172 LRLFEYNEKQFF 184
Db 355 --ARFEKFGQSF 365

RESULT 9
US-10-335-977-6717
; Sequence 6717, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 6717:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...352
; SEQUENCE DESCRIPTION: SEQ ID NO: 6718:

;
;
; LOCATION: (B) LOCATION 1...345
; SEQUENCE DESCRIPTION: SEQ ID NO: 6717:
US-10-335-977-6717

Query Match      10.7%; Score 117; DB 4; Length 345;
Best Local Similarity 23.3%; Pred. No. 0.0019;
Matches 45; Conservative 29; Mismatches 51; Indels 68; Gaps 9;

Qy 6 DLNIDMTVMPLQKSHLKRDLKELLSYMP-EBHICFSDFCICEEGFVLR-----DFD 56
Db 150 NTSIDILYNTPLONETSLKEELK-LAKELPINHLSAYALSIEKNTLNLEKNAKPPSSVDFD 208
Qy 57 NSIDSEKLFCALECLSGYINYEITNFALKGHESRHNKLNWELKPHLGLG-----108
Db 209 NVV-----REVLEGGPFQKQYEVSNYA-RNVQVKHNLAYWGAKDYLCGAGAVGCVA 258
Qy 109 ---LYAVSLI-----FCNDK-----NNVRLALIR 129
Db 259 NERFYAKLTIENYIKDPLKRFQVETLNKQDKLEKFLGLRCELGVLSLLDENKVKFLIE 318
Qy 130 KTGSEFVKANNHLV 142
Db 319 ENKAFIK-NNRLLI 330

RESULT 10
US-10-335-977-6718
; Sequence 6718, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 6718:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...352
; SEQUENCE DESCRIPTION: SEQ ID NO: 6718:
```


US-10-335-977-6718

Query Match 10.7%; Score 117; DB 4; Length 352;
Best Local Similarity 23.3%; Pred. No. 0.0019;
Matches 45; Conservative 29; Mismatches 51; Indels 68; Gaps 9;
Qy 6 DLNIDMTVNMPLQKSHLKRDLKELLSYMP-EHICFSDFCIEEGFVLR-----DFD 56
Db 157 NISIDLIYNTPLDNETSLKEELK-LAKELPINHL SAYALSIKNTNLEKNAKPSVD 215
Qy 57 NSIDSEKLMFCALECLSGYINYEITNFALKGHESRHNKLNWELKPHLGLG----- 108
Db 216 NVV-----REVLEGGFFKQYEVSNYA-RNYQVKHNLAYWGAKDYLCGGAGVCVA 265
Qy 109 ---LYAVSL-----FCNDK-----NNNVRALIR 129
Db 266 NERFYAKKLIENYIKDPLKRQVETLNKQDKLEKFLGLRCELGVLSLLDENKVKFLIE 325
Qy 130 KTGSFVKANNHLV 142
Db 326 ENKAFIK-NNRLI 337

RESULT 11

US-10-335-977-6719
; Sequence 6719, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 6719:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...364
; SEQUENCE DESCRIPTION: SEQ ID NO: 6719:

US-10-335-977-6719

Query Match

Best Local Similarity 10.7%; Score 117; DB 4; Length 364;
Matches 45; Conservative 29; Mismatches 51; Indels 68; Gaps 9;
Qy 6 DLNIDMTVNMPLQKSHLKRDLKELLSYMP-EHICFSDFCIEEGFVLR-----DFD 56
Db 169 NISIDLIYNTPLDNETSLKEELK-LAKELPINHL SAYALSIKNTNLEKNAKPSVD 227
Qy 57 NSIDSEKLMFCALECLSGYINYEITNFALKGHESRHNKLNWELKPHLGLG----- 108
Db 228 NVV-----REVLEGGFFKQYEVSNYA-RNYQVKHNLAYWGAKDYLCGGAGVCVA 277
Qy 109 ---LYAVSL-----FCNDK-----NNNVRALIR 129
Db 278 NERFYAKKLIENYIKDPLKRQVETLNKQDKLEKFLGLRCELGVLSLLDENKVKFLIE 337
Qy 130 KTGSFVKANNHLV 142
Db 338 ENKAFIK-NNRLI 349

RESULT 12

US-10-425-115-261185
; Sequence 261185, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 261185
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_169815C.1.pep
US-10-425-115-261185

Query Match

Best Local Similarity 9.6%; Score 104.5; DB 4; Length 493;
Matches 31; Conservative 25; Mismatches 56; Indels 13; Gaps 3;
Qy 8 NIDMTVNMPLQKSHLKRDLKELLSYMP-EHICFSDFCIEE-----EGFVLRDFD 62
Db 240 SMDLISSLPNQTEEMWEEELRCTVDARPTHVSVVDLQIEQGTGFGQWYTPGVFPLPSDTE 299
Qy 63 KLMF--CALECLSGYINYEITNFALKGHESRHNKLNWELKPHLGLGLYAVSLLFCNDK 120
Db 300 SANFYRIASKRLSEAGYNHYEISSYCKPGYECKHNLTYNQNRPFYAFGLGSASYI----- 354
Qy 121 NNNVR 125
Db 355 -NGVR 358

RESULT 13

US-10-289-762-396
; Sequence 396, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 84.2496 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-12
Perfect score: 1981
Sequence: 1 MRVDLPLVLSLYINLSFC.....LVKIHNFNDNPFVKVLRLP 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1981	100.0	377	8	ADU98753
2	1090	55.0	204	8	ADU98751
3	863	43.6	176	2	AAV20034
4	763	38.5	156	2	AAV20035
5	258.5	13.0	379	5	ABBS4455
6	255.5	12.9	408	5	ABP27526
7	250.5	12.6	376	6	ABU01854
8	250.5	12.6	376	8	ADK46982
9	249	12.6	376	5	ABP27525
10	248.5	12.5	409	8	ADR95677
11	248.5	12.5	409	9	AEA59547
12	244	12.3	376	8	ADV87963
13	244	12.3	376	8	ADV79216
14	244	12.3	376	8	ADV81406
15	237	12.0	382	9	AEBS3646
16	233.5	11.8	375	9	AEBA0038
17	232.5	11.7	385	6	ABM70978
18	231.5	11.7	385	6	ABJ18960
19	206.5	10.4	381	9	AEA49154
20	199	10.0	385	5	ABBA4803
21	196.5	9.9	391	5	ABP38845
22	196.5	9.9	393	7	ADC97174
23	190.5	9.6	397	7	ADH88000
24	185.5	9.4	391	6	ABP77997

25	185	9.3	377	6	ABM67734	Abm67734	Phototrhob
26	181	9.1	382	7	ADF04923	Adf04923	Bacterial
27	173	8.7	295	2	AY371169	Aay371169	Protein i
28	173	8.7	394	2	AAV34978	Aay34978	Chlamydia
29	152.5	7.7	387	7	ABO65431	Abo65431	Klebsiell
30	150.5	7.6	389	6	ADA36801	Ada36801	Acinetoba
31	144.5	7.3	442	4	AA893553	Aab93553	Human pro
32	144.5	7.3	442	8	ABM81907	Abm81907	Tumour-as
33	143.5	7.2	506	7	ABO74320	Abm74320	Pseudomon
34	142.5	7.2	401	9	ABM95708	Abm95708	M. xanthu
35	142	7.2	460	2	AAV35541	Abo35541	C. pneumo
36	140.5	7.1	424	7	ABO71949	Abo71949	Pseudomon
37	139.5	7.0	1182	3	AA818288	Aab18288	Plasmodiu
38	135.5	6.8	819	6	ABU19076	Abu19076	Protein e
39	135	6.8	320	2	AAV37405	Aay37405	Protein i
40	128.5	6.5	442	4	AA895029	Abd95029	Human pro
41	128	6.5	479	8	ADL04508	Adl04508	M. catarr
42	124.5	6.3	548	8	ADS08193	Ads08193	Staphyloc
43	124.5	6.3	605	6	ABU26712	Abu26712	Protein e
44	124	6.3	473	6	ABP80789	Abp80789	N. gonorr
45	123.5	6.2	4987	2	AAR10834	Aar10834	Rianodin

ALIGNMENTS

RESULT 1
ADU98753
ID ADU98753 standard; protein; 377 AA.
AC ADU98753;
XX
XX
DT 24-FEB-2005 (first entry)
DE Borrelia burgdorferi antigenic polypeptide seqid 12.
XX
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
XX DNA library.
XX
OS Borrelia burgdorferi.
XX
XX
PN WQ2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
XX
DR WPI; 2004-834155/82.
DR N-PSDB; ADU98752.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises
PT providing to the subject at least one Borrelia antigen or its fragment.
XX
XX
PS Claim 27; SEQ ID NO 12; 121pp; English.
XX
XX
CC The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide
CC encoding a polypeptide for an ability to produce an immune response;
CC preparing a vaccine; vaccinating a subject; treating a subject infected

XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX DR WPI; 1999-189980/16.
XX DR N-PSDB; AAX61731.
XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop
XX PT products for the diagnosis, prevention and treatment of diseases caused
XX PT by Borrelia, particularly Lyme disease.
XX PS Claim 12; Page 169; 275pp; English.
XX CC This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX CC invention, which is suitable for use in a vaccine. The Bb polypeptides
XX CC can be used in vaccines for eliciting protective antibodies to members of
XX CC the Borrelia genus, particularly for the use against Lyme disease in
XX CC humans and animals. They can be used for preventing or attenuating an
XX CC infection caused by a member of the Borrelia genus. The products can also
XX CC be used for detection of members of the Borrelia genus
XX SQ Sequence 176 AA;
Query Match 43.6%; Score 863; DB 2; Length 176;
Best Local Similarity 98.2%; Pred. No. 2.6e-69;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MRVLLPLVELSLYINLSFCCKDFSIENRILEELKCHLLILGHPIIKTYIKHVDVDFCLSR 60
DB 1 MRVLLPLVELSLYINLSFCCKDFSIENRILEELKCHLLILGHPIIKTYIKHVDVDFCLSR 60
QY 61 QDNLKFIPTSLSKYINLELLEEFTEIIPGYVDPEKFKLLDEFCITRINLNVSFSLEFR 120
DB 61 QDNLKFIPTSLSKYINLELLEEFTEIIPGYVDPEKFKLLDEFCITRINLNVSFSLEFR 120
QY 121 KIVGIPISYKMLNINIRKFFPDNIDMTNMPLOKSHLKRDLKEL 170
DB 121 KIVGIPISYKMLNINIRKFFPDNIDMTNMPLOKSHLKRDLQRI 170
RESULT 4
AAY20035
ID AAY20035 standard; protein; 156 AA.
XX AC AAY20035;
XX DT 19-JUL-1999 (first entry)
XX DE B. burgdorferi antigenic protein, t229.aa.
XX KW Antigenic protein; vaccine; Lyme disease; infection; detection.
XX OS Borrelia burgdorferi.
XX PN WO9859071-A1.
XX PD 30-DEC-1998.
XX PF 18-JUN-1998; 98WO-US012718.
XX PR 20-JUN-1997; 97US-0050359P.
XX PR 22-JUL-1997; 97US-0053344P.
XX PR 22-JUL-1997; 97US-0053377P.
XX PR 03-SEP-1997; 97US-0057483P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (MEDI-) MEDIMMUNE INC.
XX PI Choi GH, Erwin AL, Hanson MS, Lathigra R;
XX DR WPI; 1999-189980/16.
XX DR N-PSDB; AAX61732.
XX PT New isolated Borrelia burgdorferi nucleic acids - used to develop

PT products for the diagnosis, prevention and treatment of diseases caused
XX by Borrelia, particularly Lyme disease.
XX Claim 12; Page 169; 275pp; English.
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
XX invention, which is suitable for use in a vaccine. The Bb polypeptides
XX can be used in vaccines for eliciting protective antibodies to members of
XX the Borrelia genus, particularly for the use against Lyme disease in
XX humans and animals. They can be used for preventing or attenuating an
XX infection caused by a member of the Borrelia genus. The products can also
XX be used for detection of members of the Borrelia genus
XX SQ Sequence 156 AA;
Query Match 38.5%; Score 763; DB 2; Length 156;
Best Local Similarity 98.0%; Pred. No. 2.2e-60;
Matches 147; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 21 CKDFSIFNRIIEELKCHLLILGHPIIKTYIKHVDVDFCLSRQDNLKFTFTSLSKYINLELL 80
DB 1 CKDFSIFNRIIEELKCHLLILGHPIIKTYIKHVDVDFCLSRQDNLKFTFTSLSKYINLELL 60
QY 81 EETLEIIPGYVDPEKFKLLDEFCITRINLNVSFSLEFRKIVGIPISYKMLNINIRK 140
DB 61 EETLEIIPGYVDPEKFKLLDEFCITRINLNVSFSLEFRKIVGIPISYKMLNINIRK 120
QY 141 RKEPFDNIDMTNMPLOKSHLKRDLKEL 170
DB 121 RKEPFDNIDMTNMPLOKSHLKRDLQRI 150
RESULT 5
ABB54455
ID ABB54455 standard; protein; 379 AA.
XX AC ABB54455;
XX DT 29-AUG-2003 (revised)
XX DT 16-MAY-2002 (first entry)
XX DE Lactococcus lactis protein hemN.
XX KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX OS Lactococcus lactis; IL1403.
XX PN FR2807446-A1.
XX PD 12-OCT-2001.
XX PF 11-APR-2000; 2000FR-00004630.
XX PR 11-APR-2000; 2000FR-00004630.
XX PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX DR WPI; 2002-043418/06.
XX PT New nucleotide sequence useful in the identification of Lactococcus
XX PT lactis and related species.
XX PS Claim 6; SEQ ID NO 1157; 2504pp; French.
XX CC The present invention is related to a Lactococcus lactis nucleotide
XX CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
XX CC acid sequence is useful in the detection and/or amplification of nucleic
XX CC acid sequence, particularly to identify Lactococcus lactis or related
XX CC species. The proteins of the invention are useful for the biosynthesis or
XX CC biodegradation of a composition of interest. The invention helps research
XX CC in lactic bacteria, particularly useful in the production of yogurt and

OS	Streptococcus pneumoniae; type 4 strain.	
XX	WO200277021-A2.	
XX	03-OCT-2002.	
XX	27-MAR-2002; 2002WO-IB002163.	
XX	27-MAR-2001; 2001GB-00007658.	
XX	(CHIR-) CHIRON SPA.	
PA	(GENO-) INST GENOMIC RES.	
XX	Maignani V, Tettelin H, Fraser C;	
XX	WPI; 2003-040579/03.	
DR	N-PSDB; ABX07142.	
XX	New proteins and nucleic acid molecules from Streptococcus pneumoniae,	
XX	useful as medicaments for treating or preventing a disease or infection	
XX	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or	
XX	ear infection.	
XX	Claim 1; SEQ ID NO 2860; 56pp; English.	
XX	The invention relates to a protein comprising or having at least 50%	
XX	identity to any of the 2469 amino acid sequences, identified in the	
XX	specification (available on a computer readable format), or its fragment,	
XX	expressed from 2469 of 2489 identified DNA coding regions from the	
XX	Streptococcus pneumoniae type 4 strain genomic sequence appearing as	
XX	ABS56454. Also included are an antibody which binds one of the proteins,	
XX	treating a patient by administering the protein, DNA or antibody (in a	
XX	composition), a kit comprising first and second primers, which are the	
XX	nucleic acid cited above or fragments between nucleotides 8-100 of a	
XX	sequence not defined in the specification, for amplifying a target	
XX	sequence contained within a Streptococcus nucleic acid sequence, where	
XX	the first primer is substantially complementary to the target sequence	
XX	and the second primer is substantially complementary to the complement of	
XX	the target sequence, and where the parts of the primers having	
XX	substantial complementarity define the termini of the target sequence to	
XX	be amplified, assay comprising contacting a test compound with the	
XX	protein, and determining whether the test compound binds to the protein	
XX	and a Streptococcus pneumoniae bacterium, where one or more genes	
XX	encoding the proteins has been rendered inactive. The proteins, nucleic	
XX	acid molecules, antibody and compositions are useful as medicaments for	
XX	treating or preventing a disease or infection due to streptococcus	
XX	bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis	
XX	media or ear infection. They are also useful in developing vaccines,	
XX	diagnostics and antibiotics. The methods are useful for identifying	
XX	immunodominant proteins. The present sequence is one of the 2469 proteins	
XX	expressed by the identified coding regions from the genomic sequence.	
XX	Note: The sequence data for this patent did not form part of the printed	
XX	specification, but was obtained in electronic format directly from WIPO	
XX	at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to	
XX	standardise OS field)	
XX	Sequence 376 AA;	
SQ	Query Match 12.6%; Score 250.5; DB 6; Length 376;	
	Best Local Similarity 26.0%; Pred. No. 9.3e-14;	
	Matches 89; Conservative 57; Mismatches 151; Indels 45; Gaps 10;	
Qy	12 SLYINLSFCCK-----DF-----SIFNRILELCKHLLGHPIKTLTKYKHDV 55	
Db	7 SAYVHIFPCTQICYCDFSKVFIKQNPVDSYLEHLLBEFRSYDI-----EKLSTLYIGGTT 62	
Qy	56 FCLSRQDNLKFIFTSLSKYINLELLEBFTLEIIPGVYDFKFKLLDBEFCITRINLVQSF 115	
Db	63 PTALSAPQLEVLNGLTKNLDLSLEELTIANPGDLDADKIAVLKNSAVNRVSLGVTQTF 122	
Qy	116 SLERKIVGIPETISYKKNILINNIKFPDP-LNIDMTVNPLOKXSHLKRDLKELLSYM 174	
Db	123 DDKMLKIGRSHLE-KDIYENIDRLKLAGFDNISIDILYALPGQTMESQVKNVAKATGLD 181	
Qy	175 PEHICFSDFCIBEEGVFLRDF-----DNSIDSEKLMFCALECLSNYINYEITNFA 226	
Db	182 IPHMSLYSLILENHTVFMNRRGKLPKPKEELEAEAFYIIAE-LERAGFEHYEISNFS 240	
Qy	227 LKGHESRHNKLNWELKPHLGLGLYAVSLLCNDKNNVR-----ALIRKTSFVKANNHLV 282	
Db	241 KPGFESRHNLMYWDNAEYIGIGAGAGYV-----NGVRYNKHPINHYLSAVEEGNACI 294	
Qy	283 TPELLEDEFFVYHFYFIQGLGTIQGVSRLRRLRFYNEKQFF 324	
Db	295 TEDHLSQKEQMBEEMFGLRKKSQVSM-----ARFEKFGQSF 332	
RESULT 8		
ADK46982		
ID	ADK46982 standard; protein; 376 AA.	
XX	AC ADK46982;	
XX	20-MAY-2004 (first entry)	
DT	Streptococcus pneumoniae protein, Seq ID No 3497.	
XX	Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.	
XX	Streptococcus pneumoniae.	
XX	US6699703-B1.	
XX	02-MAR-2004.	
XX	26-MAY-2000; 2000US-00583110.	
XX	02-JUL-1997; 97US-0051553P.	
PR	12-MAY-1998; 98US-0085131P.	
PR	30-JUN-1998; 98US-00107433.	
XX	(GENO-) GENOME THERAPEUTICS CORP.	
XX	Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;	
PI	WPI; 2004-212399/20.	
DR	N-PSDB; ADK44321.	
XX	New nucleic acid molecules and polypeptides useful for diagnosing,	
XX	preventing and treating pathological conditions resulting from bacterial	
XX	infection, e.g. Streptococcus pneumoniae infection, and in drug	
XX	screening.	
XX	Disclosure; SEQ ID NO 3497; 301pp; English.	
XX	The invention relates to isolated Streptococcus pneumoniae nucleic acids	
XX	and polypeptides. The nucleic acids and proteins are useful for	
XX	diagnosing, preventing and treating pathological conditions resulting	
XX	from bacterial infection, such as S. pneumoniae infection. These may also	
XX	be used for drug screening procedures. The present sequence represents a	
XX	Streptococcus pneumoniae polypeptide of the invention. Note: The sequence	
XX	data for this patent did not appear in the printed specification but was	
XX	obtained in electronic format directly from USPTO at	
XX	seqdata.uspto.gov/sequence.html.	
XX	Sequence 376 AA;	
SQ	Query Match 12.6%; Score 250.5; DB 8; Length 376;	
	Best Local Similarity 26.0%; Pred. No. 9.3e-14;	
	Matches 89; Conservative 57; Mismatches 151; Indels 45; Gaps 10;	
Qy	12 SLYINLSFCCK-----DF-----SIFNRILELCKHLLGHPIKTLTKYKHDV 55	
Db	7 SAYVHIFPCTQICYCDFSKVFIKQNPVDSYLEHLLBEFRSYDI-----EKLSTLYIGGTT 62	
Qy	56 FCLSRQDNLKFIFTSLSKYINLELLEBFTLEIIPGVYDFKFKLLDBEFCITRINLVQSF 115	
Db	63 PTALSAPQLEVLNGLTKNLDLSLEELTIANPGDLDADKIAVLKNSAVNRVSLGVTQTF 122	
Qy	116 SLERKIVGIPETISYKKNILINNIKFPDP-LNIDMTVNPLOKXSHLKRDLKELLSYM 174	
Db	123 DDKMLKIGRSHLE-KDIYENIDRLKLAGFDNISIDILYALPGQTMESQVKNVAKATGLD 181	

Db 63 PTALSAPQLEVLNGLTNLSDLSVLELTITTEANPGDLADKIAVLKNSAVNRVSLGVQTF 122
 Qy 116 SLEPRKIVGIPETSYKKNILINIRKFPDP-LNIDMTVMNPLQKKSHLKRDLKELLSYM 174
 Db 123 DDKMLKKIGRSHLE-KDIYENIDRLKLAGDNISIDLIYALPGQTMEQVKNVAKAIGLD 181
 Qy 175 PEHICFSDPICEERGFVLRDF-----DNSIDSEKLFWCALECLSNGYINYEITNFA 226
 Db 182 IPHMSLSLILENHTVFNMRMRGKPLPKPELEAEEMFEVITAE-LERAGFEHYEISNFS 240
 Qy 227 LKGHESRHNKLNWELKPHLGLGLYAVSLFCNDKNNVR-----ALIRKTSFVKANHLV 282
 Db 241 KPGFESRHNLMYDWAAYYIGAGAGSYV-----NGVRYKNHGPRIHYLSAVEEGNACI 294
 Qy 283 TFELLEDLFPVYHFIOGLGTIQGVSLRALRLRPEYNEKOFF 324
 Db 295 TEDHLSQKEQMEEMFLGRKKS GSM-----ARFEKFGQSF 332

RESULT 9
 ABP27525
 ID ABP27525 standard; protein; 376 AA.
 XX AC ABP27525;
 XX DT 02-JUL-2002 (first entry)
 XX DE Streptococcus polypeptide SEQ ID NO 4226.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 group A streptococcus; Streptococcus pyogenes; antibacterial;
 anti-inflammatory; infection; vaccine; meningitis; gene therapy.
 Streptococcus agalactiae.

WO200234771-A2.
 02-MAY-2002.
 29-OCT-2001; 2001WO-GB004789.
 27-OCT-2000; 2000GB-00026333.
 24-NOV-2000; 2000GB-00028727.
 07-MAR-2001; 2001GB-00005640.

(CHIR-) CHIRON SPA.
 (GENO-) INST GENOMIC RES.
 Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;
 Tettelin H;
 WPI; 2002-352536/38.
 N-PSDB; ABN68156.

New Streptococcus protein for the treatment or prevention of infection or
 disease caused by Streptococcus bacteria, such as meningitis, and for
 detecting a compound that binds to the protein.
 Claim 1; Page 3580; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B
 streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 (Streptococcus pyogenes), comprising one of 5493 sequences (S1), given in
 the specification. The proteins have antibacterial and anti-inflammatory
 activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 antibodies that bind (I) are used in the manufacture of medicaments for
 the treatment or prevention of infection or disease caused by
 Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 Nucleic acids encoding (I) are used to detect Streptococcus in a
 biological sample. (I) is used to determine whether a compound binds to
 (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX Sequence 376 AA;
 SQ Query Match 12.6%; Score 249; DB 5; Length 376;
 Best Local Similarity 24.9%; Pred. No. 1.3e-13;
 Matches 87; Conservative 68; Mismatches 145; Indels 50; Gaps 12;

Qy 12 SLVINISPOCK-----DFS-----IFNRILEELKCHLILGHPIIKTYIKHVD 55
 Db 7 SAYVHIPFCTQICYYCFSKVFKNQVDPVDAVLAQALIREFRSYDI-----TELRLTYIGGGT 62
 Qy 56 FCLSRQDNLKFTSLSKYINLELEEFTELEIIPGVDFPEKPLLDDEFCTITRINLVQSF 115
 Db 63 PTSISAVQLDYLLTELSDRLNLTLEEFTEANPGDLTVDKIEVLQKSAVNRVSLGVQTF 122
 Qy 116 SLEFRKIVGIPETSYKKNIL--LNNIRKFPF--DLNIDMTVMNPLQKKSHLKRDLKELLS 172
 Db 123 NDKHLKRIG---RSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALS 179
 Qy 173 YMPEHICFSDPICEERGFVLRDF-----DNSIDSEKLFWCALECLSNGYINYEITN 224
 Db 180 LNIPHLSLYSLILEHHTVFNMRMRGKJLHLPTEDEAEEMFEYITISE-MERNGFHEYEISN 238
 Qy 225 FALKGHESRHNKLNWELKPHLGLGLYAVSLFCNDKNNVRALIRKTSF-----VKAN 278
 Db 239 FTKPGFESRHNIMYDWAAYYIGAGAGSYL-----DGIR--YRNRGPIQHYLKGVSSEG 290
 Qy 279 NHLVTFELLEDLFPVYHFIOGLGTIQGVSLRALRLRFEYN-EKQPFQFI 327
 Db 291 NARLSEVLKSNEMEEELFLGLRKEGVSGIKFEQKFGTSFEKRYQIV 340

RESULT 10
 ADR95677
 ID ADR95677 standard; protein; 409 AA.
 XX AC ADR95677;
 XX DT 16-DEC-2004 (first entry)
 XX DE Novel S. pneumoniae protein sequence, SEQ ID 4312.
 XX KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
 XX bacterial infection.
 XX OS Streptococcus pneumoniae.
 XX PN US6800744-B1.
 XX PD 05-OCT-2004.
 XX PF 30-JUN-1998; 98US-00107433.
 XX PR 02-JUL-1997; 97US-0051553P.
 XX PR 12-MAY-1998; 98US-0085131P.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX Doucette-Stamm LA, Bueh D;
 XX WPI; 2004-697205/68.
 XX N-PSDB; ADR93074.
 New isolated nucleic acid encoding a Streptococcus pneumoniae
 PT polypeptide, useful for diagnosing, preventing and/or treating
 PT pathological conditions resulting from the bacterial infection.
 XX Disclosure; SEQ ID NO 4312; 151pp; English.

QY 227 LKGHESRHNKLNWELKPHGLGLYAVSLFCNDKNNVR-----ALIRKTSFVKANNHLV 282
 Db 274 KPGFESHNLMYDNAEYVIGAGASGV-----NGVRYKHGPIRHYSALVEEGNACI 327

QY 283 TPELLEDFPVVHFHIOGLGTIOQVSLRALRLRFEYNKQFF 324
 Db 328 TEDHLSQEQWEEBEMFLGRKKSVM-----ARPEKFGQSF 365

RESULT 12
 ADV87963
 ID ADV87963 standard; protein; 376 AA.
 XX
 AC ADV87963;
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE Streptococcus agalactiae protein sequence, SEQ ID 357.
 XX
 KW Antibacterial; Vaccine; bacterial infection.
 XX
 OS Streptococcus agalactiae.
 XX
 XX FR2824074-A1.
 XX
 XX 31-OCT-2002.
 XX
 XX 26-APR-2001; 2001FR-00005642.
 XX
 XX 26-APR-2001; 2001FR-00005642.
 XX
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;
 XX WPI; 2004-101891/11.
 XX
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
 PT and identification of therapeutic targets.
 XX
 XX Claim 6; SEQ ID NO 357; 2687pp; French.
 XX
 XX The present invention relates to novel Streptococcus agalactiae
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
 CC agalactiae involved in the synthesis of amino acids, cell membranes,
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and
 CC phospholipid metabolism, nucleotide metabolism including purines,
 CC pyrimidines and/or nucleosides, regulatory functions, replication,
 CC transcription, translation, protein transport, adaptation to atypical
 CC conditions, sensitivity to medicines and/or analogues, functions related
 CC to transporters, biosynthesis of cofactors, prosthetic groups and
 CC transporters, cell membrane proteins and cellular machinery. (I) are
 CC useful for the detection and/or amplification of nucleic acids.
 CC Pharmaceutical composition comprising (I) or (II) are useful for
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2
 CC contains 6617 sequence whereas the present patent only contains 2344
 CC sequences.
 XX
 XX Sequence 376 AA;
 SQ

Query Match 12.3%; Score 244; DB 8; Length 376;
 Best Local Similarity 24.6%; Pred. No. 3.6e-13;
 Matches 86; Conservative 68; Mismatches 146; Indels 50; Gaps 12;

QY 12 SLYNLSFCK-----DFS-----TFNRILEELKCHILLGHPIKTYIKHVD 55
 Db 7 SAYVHIPPTQICYYCDPSKVFIRKQNPVDYALQALIREFRSYDI-----TELRTLYIGGGT 62

QY 56 FCLSRQDNLFKFTSLSKYINLELEFTEIIPGVDFEFKFLDLDFCITRINLVQSF 115
 Db 63 PTSISAVQLDYLLTELIRDNLNLTLEFTEIANPGDLTVDKIEVLQKSAVNRVSLGVQTF 122

QY 116 SLEFRKIVGPEISYKLMNI--LNNIRKPPP--DLNIDMTVNNPLOKKSHLKDRLKELS 172
 Db 123 NDKHLKRIG---RSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRNSVAKALS 179

QY 173 YNPEHICFSDFCIEBEGFVLRF-----DNSIDSEKLFCALECLSGVINYBITN 224
 Db 180 LNIPHLUSLYSLILEHHTVFMKVRKGLHLPTEDLEAMEFYIISE-WERNGFHEYSIN 238

QY 225 FALKGHESRHNKLNWELKPHGLGLYAVSLFCNDKNNVRALIRKTSF-----VKAN 278
 Db 239 FTKPGFESHNLMYDNAEYVIGAGASGV-----DGIR--YRNRGPIQHVLKGVSEG 290

QY 279 NHIWTFELLEDFPVVHFHIOGLGTIOQVSLRALRLRFEYN--EKQFFQFI 327
 Db 291 NARLSEVLKSNEMMEELFLGLRKEGVSIGKFEQKFGTSFEKRYQIV 340

RESULT 13
 ADV79216
 ID ADV79216 standard; protein; 376 AA.
 XX
 AC ADV79216;
 XX
 DT 24-FEB-2005 (first entry)
 XX
 DE Streptococcus agalactiae protein, SEQ ID 357.
 XX
 KW Antibacterial; vaccine; bacterial infection.
 XX
 OS Streptococcus agalactiae.
 XX
 XX WO200292818-A2.
 XX
 XX 21-NOV-2002.
 XX
 XX 26-APR-2002; 2002WO-IB003059.
 XX
 XX 26-APR-2001; 2001FR-00005642.
 XX
 XX (INSP) INST PASTEUR.
 XX (CNRS) CNRS CENT NAT RECH SCI.
 XX
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
 XX WPI; 2004-101891/11.
 XX
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
 PT and identification of therapeutic targets.
 XX
 XX Claim 6; SEQ ID NO 357; 439pp; French.
 XX
 XX The present invention relates to novel Streptococcus agalactiae
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the
 CC synthesis of amino acids, cell membranes, intermediate (central)
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,
 CC regulatory functions, replication, transcription, translation, protein
 CC transport, adaptation to atypical conditions, sensitivity to medicines
 CC and/or analogues, functions related to transporters, biosynthesis of
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and
 CC cellular machinery. (I) are useful for the detection and/or amplification
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
 CC useful for treatment of a bacterial S. agalactiae infection. The complete
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The
 CC present patent is an equivalent for the basic patent FR2824074A1, which

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 14.2023 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-12
Perfect score: 1981
Sequence: 1 MRVDLLPLVLSLYINLSFC.....LVKIINHFNDFPKVKRLRP 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1981	100.0	377	2 G70181	oxygen-independent
2	258.5	13.0	379	2 C86767	oxygen-independent
3	253	12.8	396	2 AD2425	coproporphyrinogen
4	251	12.7	374	2 G97057	coproporphyrinogen
5	250.5	12.6	376	2 A98030	coproporphyrinogen
6	250.5	12.6	376	2 B95164	hypothetical prote
7	249	12.6	385	2 G83817	coproporphyrinogen
8	236	11.9	374	2 D72288	hypothetical prote
9	232	11.7	374	2 F89939	hypothetical prote
10	222.5	11.2	370	2 H70423	oxygen-independent
11	202.5	10.2	366	1 B69640	coproporphyrinogen
12	201.5	10.2	393	2 D71354	probable oxygen-in
13	199	10.0	385	2 AD1259	coproporphyrinogen
14	199	10.0	385	2 AH1621	coproporphyrinogen
15	193	9.7	399	2 C71728	probable oxygen-in
16	190.5	9.6	376	2 AB0116	probable oxygen-in
17	190	9.6	400	2 C97727	hypothetical prote
18	189	9.5	378	2 B61564	probable porphyrin
19	184.5	9.3	399	2 C81932	probable porphyrin
20	181.5	9.2	376	2 B84994	hypothetical prote
21	180.5	9.1	354	2 F82911	oxygen-independent
22	180.5	9.1	391	2 B81172	oxygen-independent
23	179.5	9.1	391	2 G62320	probable oxygen-in
24	178	9.0	383	1 B64070	coproporphyrinogen
25	169.5	8.6	412	1 S75358	coproporphyrinogen
26	169	8.5	378	2 AH0878	probable oxygen-in
27	167.5	8.5	378	2 G91107	probable oxidase {
28	166.5	8.4	373	2 B86538	coproporphyrinogen
29	166.5	8.4	373	2 A72086	coproporphyrinogen

30	166.5	8.4	378	2 B65081	hypothetical prote
31	165	8.3	345	2 E71843	oxygen-independent
32	164	8.3	352	1 B64673	coproporphyrinogen
33	164	8.3	383	2 B82673	conserved hypothet
34	163.5	8.3	378	2 B85953	probable oxidase y
35	156	7.9	457	2 E81739	oxygen-independent
36	155.5	7.8	388	2 B75557	probable coproporp
37	152	7.7	355	2 C81405	probable oxidoredu
38	149	7.5	457	2 B71476	probable coproporp
39	148.5	7.5	399	2 AE2616	hypothetical prote
40	148.5	7.5	402	2 E97398	probable oxygen-in
41	145.5	7.3	949	2 F90086	chromosomal region
42	143.5	7.2	460	2 A83454	oxygen-independent
43	142	7.2	458	2 E72023	oxygen-independent
44	142	7.2	458	2 G86601	coproporphyrinogen
45	139.5	7.0	384	2 A83597	probable oxidase p

ALIGNMENTS

RESULT 1

G70181

oxygen-independent coproporphyrinogen III oxidase homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: G70181

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997

A;Authors: Smith, H.O.; Venter, J.C.

A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A;Reference number: A70100; MUID:98065943; PMID:9403685

A;Accession: G70181

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-377 <KLE>

A;Cross-references: UNIPROT:O51600; UNIPARC:UPI0000057576; GB:AE001167; GB:AE000783; NID

A;Experimental source: strain B31

Query Match	100.0%	Score	1981;	DB	2;	Length	377;
Best Local Similarity	100.0%	Pred. No.	1.4e-128;				
Matches	377;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MRVDLLPLVLSLYINLSFCCKDFSI	FNRI	LEELKCHLIL	LGHPIIKTL	YIKHVD	FCLSR 60
Db	1	MRVDLLPLVLSLYINLSFCCKDFSI	FNRI	LEELKCHLIL	LGHPIIKTL	YIKHVD	FCLSR 60
Qy	61	QDNLKFTFTSLSKYINLELEEF	TLEII	PGYVDFE	KPKLDE	FCITRIN	LNVSQFSLEFR 120
Db	61	QDNLKFTFTSLSKYINLELEEF	TLEII	PGYVDFE	KPKLDE	FCITRIN	LNVSQFSLEFR 120
Qy	121	KIVGIPETISYKKNILINIRK	PPDL	INDMTVM	PLQKSHL	KRDLK	ELLSYMPHEICF 180
Db	121	KIVGIPETISYKKNILINIRK	PPDL	INDMTVM	PLQKSHL	KRDLK	ELLSYMPHEICF 180
Qy	181	SDPICEEGFVLDFD	NSID	SEKLF	WFCAL	ECLE	SGYINYEITNEALKGHESRHNKLWE 240
Db	181	SDPICEEGFVLDFD	NSID	SEKLF	WFCAL	ECLE	SGYINYEITNEALKGHESRHNKLWE 240
Qy	241	LKPHLGLGYAVSLL	FCND	KNNVR	ALIRK	TG	SVFVKANNHLVTFFELLEDLEFFVYHFIQG 300
Db	241	LKPHLGLGYAVSLL	FCND	KNNVR	ALIRK	TG	SVFVKANNHLVTFFELLEDLEFFVYHFIQG 300
Qy	301	LGTIQGVSRALRLREYNEK	QFQ	QINVC	STISK	VP	FDNIMLLKGRFRKFNFLVK 360
Db	301	LGTIQGVSRALRLREYNEK	QFQ	QINVC	STISK	VP	FDNIMLLKGRFRKFNFLVK 360
Qy	361	IINHFNDFPKVKRLRP	377				
Db	361	IINHFNDFPKVKRLRP	377				

RESULT 2
C86767
oxygen-independent coproporphyrinogen III oxidase [imported] - Lactococcus lactis subsp. 1
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86767
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86767
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-379 <STO>
A;Cross-references: UNIPROT:Q9CGF7; UNIPARC:UPI000000C69B3; GB:AE005176; PID:gl2724101; F
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: hemN
C;Superfamily: oxygen-independent coproporphyrinogen oxidase
Query Match 13.0%; Score 258.5; DB 2; Length 379;
Best Local Similarity 22.8%; Pred. No. 1.6e-10;
Matches 91; Conservative 78; Mismatches 133; Indels 97; Gaps 15;
Qy 12 SLXINLSFCK- ----DFSIFNRILELKHLLILGHPI-----IKT 48
Db 7 SAYPHIFPCSHICYCDFA-----KVLMTGQPIDAYIBSLIEEFQSPFIEIKRT 55
Qy 49 LYIKHVPDCLSRQDNLFIFTSLSKYINLELLEFTEIIPGYVDPEKFKLLDBFCITRI 108
Db 56 IYIGGGTPSVLSAQQLERLLTAIAEQLDLEVLVEFTVEANPGDLSDEVIKVLADSAVNRI 115
Qy 109 NLNVQSFSLPRKIVGPIETSYKKLNILINNIRKFPD-LNIDMTVMPLQKSHLKRDL 167
Db 116 SLGVQTFNNALLKIGRTHTEVQVYD-SVERLKKAGFENITIDLIYALPGQTMEMVSDV 174
Qy 168 KELLSYMPEHICFSDFICEE-----EGFV-LRQFDNSIDSEKLFWCALECLENGY 217
Db 175 EKFLLELPHVALYSLILEDDHTVFMNQRRGLLRLPSDEKNAD---MYEIMDLAKNGY 231
Qy 218 INYEITTFALKGHESRNKLNWELKPHLGLGLYAVSLLFCNDKNNVRALIRKTGSFVKA 277
Db 232 NHYEVSFGLPGFESKHNITWDNEEYVIGAGASGYL-----AGIR----- 273
Qy 278 NNHLVTELEDDLEFFVHFQGLTIOGVSRLRALRLRFEYNEKQPFQFNYCSTLSKPF 337
Db 274 -----YKNLGP-----VHHYLKAAPTKRIINEEVLKSKSQIEEMFL-----GLRKK- 315
Qy 338 VFDDNIMLLKGRERFKLNF-----YLVKLIINH-----FND 367
Db 316 ---SGVLVEKPEKPFKCSFEKLYGEQITELINQKLLYND 351
RESULT 3
AD2425
coproporphyrinogen III oxidase [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AD2425
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AD2425
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <KUR>
A;Cross-references: UNIPROT:Q8YMH7; UNIPARC:UPI000000CECD; GB:BA000019; PIDN:BA076655.1;
A;Experimental source: strain PCC 7120
C;Genetics:

A;Gene: all4956
C;Superfamily: oxygen-independent coproporphyrinogen oxidase
Query Match 12.8%; Score 253; DB 2; Length 396;
Best Local Similarity 25.2%; Pred. No. 4e-10;
Matches 91; Conservative 65; Mismatches 159; Indels 46; Gaps 13;
Qy 6 LPVELSLVINLSFCK- ----DFSIF--NRILEE-----LKCHLILL-----GH 43
Db 6 VPSLASSAVHIIPFCRRRCFYCDPIFVVDGDRGRTSVTISGYVDVLCIEIAITPAFGQ 65
Qy 44 PIITLYIKHVPDCLSRQDNLFIFTSLSKYINLELLEFTEIIPGYVDPEKFKLLDBF 103
Db 66 P-LKTVFFGGGTPSLLSTEQAQILVTLQKQFIAPDAEISMEVDGTFDLAHIQGYRSV 124
Qy 104 CITRINLVQSFSLPRKIVGPIETSYKKLNILINNIRKFPF-DLNIDMTVMPLQKSH 162
Db 125 GNVRSVLGVQAQOEELLKVAGRSH-SLKQIFAADLIHQVEIPEFSIDLSGLPHQSLDQ 183
Qy 163 LKRDLKELLSYMPHEHICFSDFICEE---GFVLRDPDNSIDSE---KMFCALECLESN 215
Db 184 WQSLDTAVNIATHISYIDLTIETPGTAFGRYKPGDNPLPTDETTVMYQMGOKILTGG 243
Qy 216 GYINYEITTFALKGHESRNKLNWELKPHLGLGLYAVSLLFCNDKNNVRALIRKTGSFV 275
Db 244 DYEHYEISYAKPGHQCRHNRVYWNRPYYGFCMGAAASYV---EGKRFTRP--RKTKEYY 298
Qy 276 K-----ANNHLVTELEDDLEFFVHFQGLTIOGVSRLRALRLRFEYNEKQPFQFNY 329
Db 299 QWYQELIANHGVIDWEITPKADVLLLETLMLGLADGVSLAALTEEF---GKEKIQELHQ 355
Qy 330 C 330
Db 356 C 356
RESULT 4
G97057
coproporphyrinogen III oxidase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: G97057
R;Nolling, J.; Berton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: G97057
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-374 <KUR>
A;Cross-references: UNIPROT:Q97JJ5; UNIPARC:UPI000000CAL57; GB:AE001437; PIDN:AAK79250.1;
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1279
C;Superfamily: oxygen-independent coproporphyrinogen oxidase
Query Match 12.7%; Score 251; DB 2; Length 374;
Best Local Similarity 26.5%; Pred. No. 5.2e-10;
Matches 93; Conservative 62; Mismatches 142; Indels 54; Gaps 15;
Qy 10 ELSIYINLSFC-----CKDFSIFNRILELKHLLILGHPIIKTLVYKHVDFCLSRQDN 63
Db 3 KGLYIHIIPCKSKLYC-DFPSYACIED-----FYAKALCTETEKASINKVFS 53
Qy 64 LKFTFTSLSYKINLELLE-----EFTLEIIPGYVDPEKFKLLDBFCITRI 108
Db 54 SIFIGGGTPSFLSTEALNLCALDKVKRKTKDVEFTVEGNPNSFTEKLMWFKDMGVNRL 113
Qy 109 NLNVQSFSLPRKIVGPIETSYKKLNILINNIRKFPF-DLNIDMTVMPLQKSHLKRDL 167
Db 114 SIGLQACQDRLKLLKLGRIH-TLKEFTVAFKRARNLGNINVDLMFGIPDQTLDFDKESL 172

```
Qy 168 KELLSPHEHICPSDFICE-----EGFVLDRDFNSIDSEKLMFCALECLSNQYIN 219
: : ||||| : : : : : : : : : : : :
Db 173 EFITKLPHEHISYSLIVEGTPYFKNQEGKLG-LPNEDEEDRMYSFARTFLEEKGYNQ 231
: : ||||| : : : : : : : : : : : :
Qy 220 YEITNPAKGHESRHNKLWELKHGILGYAVSLLECFND---KN-NVNRALIRK--TGS 273
||||| : : : : : : : : : : : : : : : :
Db 232 YEISNFAVKQKCRHNLITYWELDNYIGCGASAHG--YFNGVRVYNNNNVKKYIEQISKGN 289
||||| : : : : : : : : : : : : : : : :
Qy 274 FVKANNHLVTFELLEDLEFFVYHFHFIQGLGTIQGVSLRALRLRFEYNEKQPF 324
||||| : : : : : : : : : : : : : : : :
Db 290 SVVZENHRL--LKEDMEEFMF---LGLRKTGRGVSIEEFKLKFNKDIOEVY 335
||||| : : : : : : : : : : : : : : : :

RESULT 5
A98030
coproporphyrinogen oxidase (EC 1.3.3.3) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: A98030
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; UID:21429245; PMID:11544234
A:Accession: A98030
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <R>
A:Cross-references: UNIPROT:Q97Q22; UNIPARC:UPI0000051830; GB:AE007317;
C:Genetics:
A:Gene: hemN
C:Superfamily: oxygen-independent coproporphyrinogen oxidase
C:Keywords: oxidoreductase

Query Match 12.6%; Score 250.5; DB 2; Length 376;
Best Local Similarity 26.0%; Pred. No. 5.6e-10;
Matches 89; Conservative 57; Mismatches 151; Indels 45; Gaps 10;

Qy 12 SLYNLSFCCK-----DF-----SIFNRILEELKHLILGHPIIKTYIKHVD 55
: : ||||| : : : : : : : : : : : :
Db 7 SAYVHIPFCTQICYCDFSKVFIKNQPVDSVLEHLEEFPSYDI-----EKLSTLYIGGGT 62
||||| : : : : : : : : : : : : : : : :

Qy 56 FCLSRQNLKFIFTSLSKYINLELLEFTLEIIPGVYVDFEKFLLDFCITRINLNQSF 115
: : ||||| : : : : : : : : : : : :
Db 63 PTALSAPOLEVLNGLTKNLDLSVLELTIEANPGDLADKIAVLKNSAVNRVSLGVQTF 122
||||| : : : : : : : : : : : : : : : :

Qy 116 SLEPRKIVGPIEISYKKNILNINIRKPPD-LNIDMTVMPLQKSHLKDRLKELLSYM 174
: : ||||| : : : : : : : : : : : :
Db 123 DDKMLKKIGRSHLE-KDIYENIDRLKLAGFONISIDLIYALPGQTMEQVKENAKAIGLD 181
||||| : : : : : : : : : : : : : : : :

Qy 175 PEHICFSDFCIEEGFVLDRF-----DNSIDSEKLMFCALECLSNQYINVEITNEA 226
: : ||||| : : : : : : : : : : : :
Db 182 IPHMSLYSLIDENHTVFMNRRGKLPKPKEELEAEFXYIIAE-LERAGFEHYEISNFS 240
||||| : : : : : : : : : : : : : : : :

Qy 227 LKGHESRHNKLWELKHGILGYAVSLLECFNDKNNVR-----ALIRKTSFVKANNHLV 282
||||| : : : : : : : : : : : : : : : :
Db 241 KPGFESRHNLMYDNAEYVIGAGASGVY-----NGVRVYKNGHPIRHYLSAVEEGNACI 294
||||| : : : : : : : : : : : : : : : :

Qy 283 TFELLEDLEFPVYHFHFIQGLGTIQGVSLRALRLRFEYNEKQPF 324
: : ||||| : : : : : : : : : : : :
Db 295 TEDHLSQKEQMEEMFGLRKKSGVSM---ARFEKFGQSF 332
||||| : : : : : : : : : : : : : : : :

RESULT 7
G83817
coproporphyrinogen III oxidase hemN [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: G83817
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; UID:20512582; PMID:11058132
A:Accession: G83817
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: UNIPROT:Q9K075; UNIPARC:UPI000000C3B7C; GB:AF001511; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: hemN
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.6%; Score 249; DB 2; Length 385;
Best Local Similarity 25.5%; Pred. No. 7.3e-10;
Matches 100; Conservative 60; Mismatches 156; Indels 76; Gaps 17;

Qy 14 YINLSFC-----CKDPSIF-----NRILEELKCH--LILLGHPI--IKTYIKHVD 56
||||| : : ||||| : : : : : : : : : : : :
Db 6 YIHIPFCEHCYCYC-DPNKFKYLNQPVNVEYLQALETEMAMVVAEQPKYSQTLTVGGGTP 64
||||| : : ||||| : : : : : : : : : : : :
Qy 57 CLSRQNLKFIFTSLSKYINLELLEFTLEIIPGVYVDFEKFLLDFCITRINLNQSF 116
||||| : : ||||| : : : : : : : : : : : :
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
```

```
nsen, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; UID:21357209; PMID:11463916
A:Accession: B95164
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <R>
A:Cross-references: UNIPROT:Q97Q22; UNIPARC:UPI0000051830; GB:AE005672;
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI409
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.6%; Score 250.5; DB 2; Length 376;
Best Local Similarity 26.0%; Pred. No. 5.6e-10;
Matches 89; Conservative 57; Mismatches 151; Indels 45; Gaps 10;

Qy 12 SLYNLSFCCK-----DF-----SIFNRILEELKHLILGHPIIKTYIKHVD 55
: : ||||| : : : : : : : : : : : :
Db 7 SAYVHIPFCTQICYCDFSKVFIKNQPVDSVLEHLEEFPSYDI-----EKLSTLYIGGGT 62
||||| : : : : : : : : : : : : : : : :

Qy 56 FCLSRQNLKFIFTSLSKYINLELLEFTLEIIPGVYVDFEKFLLDFCITRINLNQSF 115
: : ||||| : : : : : : : : : : : :
Db 63 PTALSAPOLEVLNGLTKNLDLSVLELTIEANPGDLADKIAVLKNSAVNRVSLGVQTF 122
||||| : : : : : : : : : : : : : : : :

Qy 116 SLEPRKIVGPIEISYKKNILNINIRKPPD-LNIDMTVMPLQKSHLKDRLKELLSYM 174
: : ||||| : : : : : : : : : : : :
Db 123 DDKMLKKIGRSHLE-KDIYENIDRLKLAGFONISIDLIYALPGQTMEQVKENAKAIGLD 181
||||| : : : : : : : : : : : : : : : :

Qy 175 PEHICFSDFCIEEGFVLDRF-----DNSIDSEKLMFCALECLSNQYINVEITNEA 226
: : ||||| : : : : : : : : : : : :
Db 182 IPHMSLYSLIDENHTVFMNRRGKLPKPKEELEAEFXYIIAE-LERAGFEHYEISNFS 240
||||| : : : : : : : : : : : : : : : :

Qy 227 LKGHESRHNKLWELKHGILGYAVSLLECFNDKNNVR-----ALIRKTSFVKANNHLV 282
||||| : : : : : : : : : : : : : : : :
Db 241 KPGFESRHNLMYDNAEYVIGAGASGVY-----NGVRVYKNGHPIRHYLSAVEEGNACI 294
||||| : : : : : : : : : : : : : : : :

Qy 283 TFELLEDLEFPVYHFHFIQGLGTIQGVSLRALRLRFEYNEKQPF 324
: : ||||| : : : : : : : : : : : :
Db 295 TEDHLSQKEQMEEMFGLRKKSGVSM---ARFEKFGQSF 332
||||| : : : : : : : : : : : : : : : :

RESULT 7
G83817
coproporphyrinogen III oxidase hemN [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: G83817
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; UID:20512582; PMID:11058132
A:Accession: G83817
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-385 <STO>
A:Cross-references: UNIPROT:Q9K075; UNIPARC:UPI000000C3B7C; GB:AF001511; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: hemN
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 12.6%; Score 249; DB 2; Length 385;
Best Local Similarity 25.5%; Pred. No. 7.3e-10;
Matches 100; Conservative 60; Mismatches 156; Indels 76; Gaps 17;

Qy 14 YINLSFC-----CKDPSIF-----NRILEELKCH--LILLGHPI--IKTYIKHVD 56
||||| : : ||||| : : : : : : : : : : : :
Db 6 YIHIPFCEHCYCYC-DPNKFKYLNQPVNVEYLQALETEMAMVVAEQPKYSQTLTVGGGTP 64
||||| : : ||||| : : : : : : : : : : : :
Qy 57 CLSRQNLKFIFTSLSKYINLELLEFTLEIIPGVYVDFEKFLLDFCITRINLNQSF 116
||||| : : ||||| : : : : : : : : : : : :
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
```

```
Db 65 TALTAQOLASIKRTPLSLDEFTFVNPDSIDEKLDVLSYGVDRSLSIGVQAFQ 124
Qy 117 LEPRKIVGIPBISYKKNILINIRKPPF-DLNIDMTVNNPQKSKHKLKDLKELLSYMP 175
Db 125 PLLLKETIGRTH-DQKSVEQAVEKSRQAGFANLSLDMLGLPKQTPPEMFAETLKEAFALV 183
Qy 176 EHI-CFSDFICEEGFVLDPDNGI-----DSKLFWCALECLSGNYINYEITNFALK 228
Db 184 EHLSCYSLKVEAKTVFNRRQGRRLTLPPEDEVMKTRQLCYETEKHGFQYELSNPAKK 243
Qy 229 GHESRHNKLNWELKPHLGLGL-----YAVSLFCNDKNNVRALIRKTGSFVK----- 276
Db 244 GYESRNLVYNNDEYVGFAGAGHYGVGVYNNH-----GPLPKYLQAMEEG 291
Qy 277 -----ANNHLVTFELLEDLEFFVHFIOGLGTIOGVSRLRALRPFYN-----EKQFPQF 326
Db 292 RRPVFESHVSRVQBEMQEL-----GLKRSQVBEVRFVVERFGVSMFSLYEKQIAQL 345
Qy 327 INYCSLTKKVFDDNIM-----LLKGRERFK 353
Db 346 VARC-LLERT---DDRVLTDGLLGNVEFVE 373

RESULT 8
D72288
hypothetical protein TM1166 - Thermotoga maritima (strain MS88)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: D72288
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: D72288
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <ARN>
A:Cross-references: UNIPROT:Q9XOP9; UNIPARC:UPI00000D3914; GB:AE001773; GB:AE000512; NID
A:Experimental source: strain MS88
C:Genetics:
A:Gene: TM1166
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 11.9%; Score 236; DB 2; Length 374;
Best Local Similarity 22.6%; Pred. No. 5.5e-09;
Matches 89; Conservative 68; Mismatches 148; Indels 88; Gaps 12;
Qy 9 VELSLYNLSFC-----CKDPSI-----FNRILBELKCHLILLGHPIIKTYIKHV 54
Db 1 MKLAVYHVHPFCKSKVCYVSVSEKDFYFSHLLREIDLYBEVLSESEIKTVYFGG 60
Qy 55 DFCLSRQDNLAFIFTSLSKYINLELLEFTLEIIPGYVDFEKFLLDFCITRINLVQS 114
Db 61 TPSVPPPSFLKWLKLEKRVSRGFTPDDEITIEVPESVETELKLYKQIGNRLSLGVQA 120
Qy 115 FSLEFRKIVG---IPEISYKKNILINIRKPPFDLNIDMTVNNPQKSKHKLKDLKELL 171
Db 121 CDDTVLKNAGRLYKEETLKKAKIVLEQFE---NVNFDLILGLPGETDITLKKDFRLE 176
Qy 172 SYMEHICFSPDICE-----EGFVLRDPDNSIDSEKLFWCALECLSGNYINY 220
Db 177 BFPQGHV--SLYLLVEDRTRLDLIQKGLV--ELPEDDVVERRHDLFVEFLKRGFLRY 232
Qy 221 EITNFALKGHESRHNKLNWELKPHLGLGLYAVSLFCNDKNNVRALIRKTGVSFKVANNH 280
Db 233 EISNPAKCKSKLMLFWRNENYLGIG-----VSAGGH 266
Qy 281 LVTFELL--EDLEFF-----VYHFI-----OGLGTIOGVSRLRALRLR 315
Db 267 IGRFRYNASDLKEYEEKITKGLPEYVYVHENTEDEEALETVFMGLRIKGVNELNRVKIL 326
```

```
Qy 316 FEYNEKQFFOFINYCSTLSKK-FVFDNIMMLK 347
Db 327 LPLEKLQKYPCLYKVKNGKIFLSEDDGMFSK 359
```

RESULT 9

F89939
hypothetical protein hemN [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: F89939
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguci
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89939
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-374 <KUR>
A:Cross-references: UNIPROT:Q99TR6; UNIPARC:UPI00000D767E; GB:BA000018; PID:gl3701381; P
A:Experimental source: strain N315
C:Genetics:
A:Gene: hemN
C:Superfamily: oxygen-independent coproporphyrinogen oxidase

Query Match 11.7%; Score 232; DB 2; Length 374;
Best Local Similarity 25.9%; Pred. No. 1e-08;
Matches 101; Conservative 73; Mismatches 142; Indels 74; Gaps 19;

```
Qy 12 SLYNLSFPCCK--DFSIFNRILBELKCHLILLGHPIIKTYIKHVDFCLSRQDNLKPIF- 68
Db 5 SAVIHIPFCVRICTYCDFNK-----YFIQNPQVDEYLDALITENSTAKYRNKLKTMIV 56
Qy 69 -----TSLSKYIN-LELL-----EEFTLEIIPGYVDFEKFLLDFCITRINLN 111
Db 57 GGGTPTALS--INQLERLLKAIIRDFTITGEYTFEANPDELTKKVLLEKYGVNRISM 114
Qy 112 VQSFSLEFRKIVG-----IPEISYKKNILINIRKPPFDLNIDMTVNNPQKSKHKLKRD 167
Db 115 VQTFKPELLSVLGRTHNTEDIYTSVLNKNAGIK---SISDLMYHLPKQITDFEQSL 170
Qy 168 KELLSYMPHICFSDFICE-----EGFV-LRDPDNSIDSEKLFWCALECLSGNY 217
Db 171 DLALDMDIQHISYGLILEPKTFQFNNYRKGLLKPNEIDLADMYQL---LMSKIQSPF 227
Qy 218 INYEITNFALKGHESRHNKLNWELKPHLGLGLYA-----VSLFCNDKNNVRALIRKT 271
Db 228 HQYEISNFALDGHSEHKNKYWNVEYYGFGAGASGYVDGVRVYTNINPNNHYIKAINKES 287
Qy 272 GSFVKANNHLVTFELLEDLEFFVHFIOGLGTIOGVSRLRALRPFYN-EKQFPQFINYC 330
Db 288 KAILVSNKPSLT-ERMEE-EMFL-----GLRLNEGVSRRFKKFKDQSIESVFQGTIN-- 338
Qy 331 STLSSKKVFDDN--IMLLK-----GRERFK 353
Db 339 NLKEKELIVEKNDAIALTKRGKVGIGNEVFE 369
```

RESULT 10

H70423
oxygen-independent coproporphyrinogen III oxidase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: H70423
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: H70423

[illegible]

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 81.0092 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-12
Perfect score: 1981
Sequence: 1 MRVDLLPLVELSYINLSFC.....LVKLIHFNDNFFKVKLRLP 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1981	100.0	377	2	051600 BORBU
2	1716	86.6	377	2	066014 BORBU
3	292.5	14.8	365	2	08R695 FUSNN
4	280	14.1	375	2	08RB71 THEYN
5	277	14.0	382	2	089207 CLOTE
6	275.5	13.9	377	2	08X158 CLOPE
7	265	13.4	378	2	05L2C1 STRT1
8	262	13.2	378	2	05M3X3 STRT2
9	258.5	13.0	379	2	09CGF7 LACLA
10	256.5	12.9	376	2	08P177 STRP8
11	256.5	12.9	398	2	087811 STRP3
12	256.5	12.9	408	2	08K7B5 STRP3
13	255.5	12.9	376	2	09XZM6 STRP1
14	255.5	12.9	408	2	05XCD8 STRP6
15	253	12.8	396	2	08YMH7 ANASP
16	251	12.7	374	2	097J35 CLOAB
17	250.5	12.6	376	2	08DPA6 STRR6
18	250.5	12.6	376	2	097Q22 STRPN
19	250	12.6	376	2	08E046 STRAS
20	249	12.6	385	2	09KD75 BACHD
21	244	12.3	376	2	08E583 STRA3
22	239.5	12.1	374	2	08G877 STARR
23	239.5	12.1	383	2	093R30 TETHA
24	237	12.0	375	2	05ZT11 LEGPH
25	236	11.9	374	2	09X0P9 THEMA
26	236	11.9	376	1	Y532 BUCAP
27	235	11.9	374	1	Y498 BUCBP
28	233.5	11.8	375	2	05WU92 LEGPL
29	233.5	11.8	375	2	05X2T4 LEGPA
30	232.5	11.7	380	2	08DTJ3 STRMU
31	232	11.7	374	2	06G8Y4 STAA5

32	232	11.7	374	2	Q7A0S0_STAAW	Q7A0S0 staphylococ
33	232	11.7	374	2	Q7A5C2_STAAW	Q7A5C2 staphylococ
34	232	11.7	374	2	Q99TR6_STAAW	Q99TR6 staphylococ
35	231.5	11.7	385	2	Q5HFH7_STAAC	Q5HFH7 staphylococ
36	223.5	11.3	375	2	Q83DU5_COXBU	Q83DU5 coxiella bu
37	223	11.3	388	2	Q5WHG4_BACSK	Q5WHG4 bacillus cl
38	222.5	11.2	370	2	Q67418_AQUAE	Q67418 aquifex ao
39	218.5	11.0	391	2	Q8CXD1_OCEIH	Q8CXD1 oceanobacil
40	215	10.9	379	2	Q6SH51_BACLD	Q6SH51 bacillus li
41	213.5	10.8	349	2	Q9FOM0_MYCIO	Q9FOM0 mycoplasma
42	213.5	10.8	349	2	Q9FOM1_MYCIO	Q9FOM1 mycoplasma
43	213.5	10.8	349	2	Q9FOM2_MYCIO	Q9FOM2 mycoplasma
44	213	10.8	375	2	Q8KAY8_CHLTE	Q8KAY8 chlorobium
45	212	10.7	376	2	Q8A5S0_BACTN	Q8A5S0 bacteroides

ALIGNMENTS

RESULT 1
OS1600 BORBU PRELIMINARY; PRT; 377 AA.
AC OS1600;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase, putative.
GN OrderedLocusNames=BB0656;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Winn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson S.L., Peterson J.D., Kervage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Matthey L., McDonald L.A.,
RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
RL Nature 390:580-586 (1997).
DR EMBL; AE001167; AAC67009.1; -; Genomic_DNA.
DR PIR; G70181; G70181.
DR TIGR; BB0656; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR InterPro; IPR006638; E1p3/MiAB/NiFb.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical_SAM; I.
DR SMART; SM00729; E1p3; I.
KW Complete proteome.
SQ SEQUENCE 377 AA; 44756 MW; B07859CFC7793561 CRC64;

Query Match	100.0%;	Score	1981;	DB	2;	Length	377;
Best Local Similarity	100.0%;	Pred. No.	9.1e-128;				
Matches	377;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MRVDLLPLVELSYINLSFC	KDFSI	FNRI	LEELKCHLILGHPIKTLKIVHDFCLSR	60	
Db	1	MRVDLLPLVELSYINLSFC	KDFSI	FNRI	LEELKCHLILGHPIKTLKIVHDFCLSR	60	
Qy	61	QDNLKFTFTSLSKYINLELEEF	TLRI	PGYVD	FEKFLDEFCLTRINLNVSFSLEPR	120	
Db	61	QDNLKFTFTSLSKYINLELEEF	TLRI	PGYVD	FEKFLDEFCLTRINLNVSFSLEPR	120	
Qy	121	KIVGIPISYKKUNILINNIRK	PPFD	LNIDMT	VNMPLOKSHLKRDLKELLSYMPHFICF	180	
Db	121	KIVGIPISYKKUNILINNIRK	PPFD	LNIDMT	VNMPLOKSHLKRDLKELLSYMPHFICF	180	

```

Db 121 KIVGIPETSYKKNILINIRKPPFDLNDMTVMNPLQKSHLKDRLKELLSYMPHICF 180
Qy 181 SDFICEEGVLRDFDONSIDSEKLPFCALCLESNGYINYEITNFALKGHESRHNKLWE 240
Db 181 SDFICEEGVLRDFDONSIDSEKLPFCALCLESNGYINYEITNFALKGHESRHNKLWE 240
Qy 241 LKPHLGGLYAVSLFCNDKNNNRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIQ 300
Db 241 LKPHLGGLYAVSLFCNDKNNNRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIQ 300
Qy 301 LGTIQVSLRALRFRFEYNEKQFFQFINYCSTLSKFKFVFDNDIMLLKGRERFKLNFVLVK 360
Db 301 LGTIQVSLRALRFRFEYNEKQFFQFINYCSTLSKFKFVFDNDIMLLKGRERFKLNFVLVK 360
Qy 361 IINHFNDFNFKVKLRLP 377
Db 361 IINHFNDFNFKVKLRLP 377

RESULT 2
Q660L4_BORGA
ID Q660L4_BORGA PRELIMINARY; PRT; 377 AA.
AC Q660L4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE HemN-related protein.
GN OrderedLocuNames=BG0679;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=P81;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Suehnel J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the ENBL/Genbank/DBJ databases.
DR EMBL; CP000013; AAU07507.1; -; Genomic_DNA.
DR InterPro; IPR006638; Elp3/MiaB/NifB.
DR InterPro; IPR007197; Radical SAM.
DR Pfam; PF04055; Radical SAM; I.
DR SMART; SM00729; Elp3; I.
KW Complete proteome.
SQ SEQUENCE 377 AA; 44968 MW; C8192B7F0CD24051 CRC64;

Query Match 86.6%; Score 1716; DB 2; Length 377;
Best Local Similarity 85.9%; Pred. No. 1.3e-109;
Matches 324; Conservative 27; Mismatches 26; Indels 0; Gaps 0;

Qy 1 MRVDLLPLVELSLYINLSFCCKDFSIENRILEELKCHLILGHPIIKTYIKHVDCLSR 60
Db 1 MRVDLLPLVELSLYINLSFCCKDFSIENRILEELKCHLILGHPIIKTYIKHVDCLSR 60

Qy 61 QDNLKPIFTSLSKYINILEEFTLEIIPGVYDPEKPLDDEFCTITNLNVQSFSLEFR 120
Db 61 QDSLKPIFTSLSKYINLALLEEFTLEIIPGVYDPEKPLDDEFCTITNLNVQSFSLEFR 120

Qy 121 KIVGIPETSYKKNILINIRKPPFDLNDMTVMNPLQKSHLKDRLKELLSYMPHICF 180
Db 121 KIMGIPETSYKKNILINIRKPPFDLNDMTVMNPLQKSHLKDRLKELLSYMPHICF 180

Qy 181 SDFICEEGVLRDFDONSIDSEKLPFCALCLESNGYINYEITNFALKGHESRHNKLWE 240
Db 181 SDFICEEQDLTRDFDSIDSEKLPFCALCLESNGYINYEITNFALKGHESRHNKLWE 240

Qy 241 LKPHLGGLYAVSLFCNDKNNNRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIQ 300
Db 241 LKPYLGGLYAVSLFCNDKNNNRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIQ 300

Qy 301 LGTIQVSLRALRFRFEYNEKQFFQFINYCSTLSKFKFVFDNDIMLLKGRERFKLNFVLVK 360

```

```

Db 301 LGTIQVSLRALRFRFEYNDKQFFHFINYCNSLSKFKFVFDNDIMLLKGRERFKLDFYLLK 360
Qy 361 IINHFNDFNFKVKLRLP 377
Db 361 IMNYFNDFNFKVKLRLP 377

RESULT 3
Q8R695_FUSNN
ID Q8R695_FUSNN PRELIMINARY; PRT; 365 AA.
AC Q8R695;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-).
GN OrderedLocuNames=FN0560;
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 25586;
RX DOI=10.1128/JB.184.7.2005-2018.2002;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Feinstein M., Kyrpides N.C., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE009951; AAL94756.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiaB/NifB.
DR InterPro; IPR010723; HemN C.
DR InterPro; IPR004559; HemN rel.
DR Pfam; PF06969; HemN C; 1.
DR Pfam; PF04055; Radical SAM; 1.
DR SMART; SM00729; Elp3; I.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome; Oxidoreductase.
SQ SEQUENCE 365 AA; 43029 MW; 41EBF343A04309D6 CRC64;

Query Match 14.8%; Score 292.5; DB 2; Length 365;
Best Local Similarity 24.9%; Pred. No. 5.2e-12;
Matches 101; Conservative 74; Mismatches 133; Indels 97; Gaps 16;

Qy 14 YINLSPFC--CK--DFS-----IFNRILEELKCHLILGHPIIKTYIKHVDFC 57
Db 8 YIHIPFCERKCNVCDFTSLKGTSDSQEYKINYLKKEIE-----IYSKEYDLS 54

Qy 58 LSRQDNLKF-----IFTSLSKYI---NLELEEFTLEIIPGVYDPEKPLDDEFCTIT 106
Db 55 -EKQDTIYFGGTPSLLPINSLEKLSKFSYDKNTEITIEVNPKTVDTNKLKEYRKLGIN 113

Qy 107 RINLVQSFSLEFRKIVGIPETSYKKNILINIRKPPFD-LNDMTVMNPLQKSHLKR 165
Db 114 RLSIGIGITFDNDNLKVLGRHSHSQEAIEV-YNLARECGFDNISLDIMFSLPYQTLMLQN 172

Qy 166 DLKELLSYMPHEHICFSDFCIEEGFVLDRF-----DNSIDSEKLPFCALCLESNGY 217
Db 173 DLKELLSYMPHEHICFSDFCIEEGFVLDRF-----DNSIDSEKLPFCALCLESNGY 217

Qy 218 INYEITNFALKGHESRHNKLNLWELKPHLGGLYAVSL-FCNDKNNNRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIQ 360
Db 232 IHVEISNFSKDFESRHNSTYWNENKYLGVGLSAGYLNVRKYNFFNLKDYNNLDRNI 291

```

Qy 265 RALIRKTSFVKANNHLVTFELLEDLEFFVHFHTQGLTQGVSLRALRLRPFVNEKQFF 324
Db 292 LPIDEK-----EILTEEDIEQRYLVGFRLNKLIIIP-----SEK--- 326
Qy 325 QFINVCSTLSKKFVFDNIMLLKGRFKNFYLVKLIINHFNDFN 369
Db 327 -YLEKCMSLCKEG-----YLLEKENGYLSHKGLMLFNDFISNF 364
RESULT 4
Q8RB71_THETN PRELIMINARY; PRT; 375 AA.
AC Q8RB71_2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Coproporphyrinogen III oxidase and related FeS oxidoreductases.
GN Name=HemN; OrderedLocNames=TTE0952;
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336; DOI=10.1101/gr.219302;
RA Bao Q., Tian Y., Li W., Lai X., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RA "A complete sequence of the T. tengcongensis genome.";
RT Genome Res. 12:689-700(2002).
RL ENBL; AEO13060; AAM24208.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR002453; Beta tubulin.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
DR PROSITE; PS00228; TUBULIN_E_AUTOREG; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 375 AA; 44547 MW; A42BE382DBCFF122C CRC64;
Query Match 14.1%; Score 280; DB 2; Length 375;
Best Local Similarity 24.9%; Pred. No. 3.8e-11;
Matches 96; Conservative 72; Mismatches 149; Indels 68; Gaps 13;
Qy 10 ELSLYINLSFCCK-----DFSIF-----NRILEELKCH-----LILLGHPI 45
Db 3 EIGYIHPFCCKCYCYDFNSYAGYHELMEDYLKALLEEVKTYSDRSFRVISVIGGGT 62
Qy 46 IKTLYIKHVDCLSRQDNLFIFTSLSKYINLELLEFTLEIIPGYVDPEKFLDLDFCI 105
Db 63 PNFLPPSHVERVLS-----EVHKDYNVSRDAETITVNPGLLTEDKLIYKINGI 112
Qy 106 TRINLVNQSFLEFRKIVGIPISYKLNILINN---IRKPPDLNIDMTVMPLQKSH 162
Db 113 NRISMGLOAFQNRLLVEIG-----RIHTAEDPLENVALVRKYFDNVNDLIYALPTQSFEE 168
Qy 163 LKRLKLLSYMPHEICFSDPICEEGFVLDF--DNSI-----DSEKLWFCALCELESN 215
Db 169 WQETLTQVKVELKPEHISTYSLILENTLFGLYKENRLPVVGESEELKMWHTGFLKSK 228
Qy 216 GYINYEITNFALKGHESRHNKNWELPHGLGLYAVSLFLCNDKNNNVRALIRKTSFV 275

Db 229 GYCHYEISNVALGYQCRHNLWYECREYLGFGAGAHs--YFEGTRMNVNRIEKYIAIL 287
Qy 276 K---ANNHLVTFELLEDLEFFVHFHTQGLTQGVSLRALRLRPFVNEKQFF--FINY- 329
Db 288 KRKDAREEIIINLSPEDKMSFMP---LGLRMKGVCEEFRKRGISMFERYEIEFIKYE 344
Qy 330 -----CSTLSKKFVFDNIM 344
Db 345 KMGLIEKDKDCVRLTEKGDVSNVI 369
RESULT 5
Q892Q7_CLOTE PRELIMINARY; PRT; 382 AA.
AC Q892Q7_2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Oxygen-independent coproporphyrinogen III oxidase (EC 1.-.-.-).
GN OrderedLocNames=CTC02034;
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129; DOI=10.1073/pnas.0335851100;
RA Brueggemann H., Baumer S., Fricke W.F., Wierzer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,
RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
RT tetanus disease.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321(2003).
DR ENBL; AEO15943; AAO36537.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
DR Complete proteome; Oxidoreductase.
SQ SEQUENCE 382 AA; 45516 MW; 52FA3CD797A6FFAA CRC64;
Query Match 14.0%; Score 277; DB 2; Length 382;
Best Local Similarity 26.7%; Pred. No. 6.3e-11;
Matches 108; Conservative 70; Mismatches 146; Indels 80; Gaps 18;
Qy 1 MRYDLLPLVELSYINLSFC---CK--DFSIFNRILLEELKCHLILLGHPI-----IK 47
Db 1 MKEDYNNFNKNGIYIHPCKECKKCYDFISYKSKSFMEKYIEALSKEIKRECMYKIR 60
Qy 48 TLYI-----KHVDCLSRQDNLFIFTSLSKYINLELLEFTLEIIPGYVDPEK 97
Db 61 SIYIGGTPSPVKNYWEI-----LKETIDNLSKNL-----EFSIANPGVTVEKL 109
Qy 98 KLDFECITRINLVNQSFs-----LEPRKIVGIPISYK-KLNILINNIRKPPD-LNIDM 151
Db 110 KPFKEGINRISFGLQAWQHHLNKLGRHINIBEFLYSYKL-----ARKLGFNDINVDL 163
Qy 152 TVNMPQLKSHLKDCLKSLLYMPHEI-CFSDPICE-----EEGFLRDPNSIDSE 202
Db 164 MFGIPDQKLNWKEWTELEKIVELNPEHISYSLIVEGTPPYKLYERKGL--NLPNEDIER 221

Qy	229	GHESRHKNLWELKPHGLGLGYAVSLFLCNDKNNVRALIRKTG---SFVKA-----	277
Db	243	GHESRHKNLWELKPHGLGLGYAVSLFLCNDKNNVRALIRKTG---SFVKA-----	294
Qy	278	--NNHLVTFLELLEDFPVVHFIOGLGTIOGVSLRALRLEFVNEKQFFQFQINVCSTLSK	335
Db	295	LNEEHLTKVEMEE-EFFL-----GLRKSGVSIK-----RFE--EKFGLSFSDTYGDIVK	342
Qy	336	KFVFDNIMLLKGRE	350
Db	343	KQEDG--LLVKDPE	355
RESULT 8			
ID	Q5M3X3	STR2 PRELIMINARY; PRT; 378 AA.	
AC	Q5M3X3		
DT	01-FEB-2005	(TremBLrel. 29, Created)	
DT	01-FEB-2005	(TremBLrel. 29, Last sequence update)	
DE	01-FEB-2005	(TremBLrel. 29, Last annotation update)	
GN	Name=hemN; OrderedLocusNames=stui236;		
OS	Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
OX	NCBI_TaxID=264199;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].		
EX	PubMed=15543133; DOI=10.1038/nbt1034;		
RA	Bolotin A., Quinguis B., Renault P., Sorokin A., Ehrlich S.D.,		
RA	Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,		
RA	Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,		
RA	Ngui K., Masuy D., Hancy F., Burtreau S., Boutry M., Delcour J.,		
RA	Goffeau A., Hols P.		
RT	"Complete sequence and comparative genome analysis of the dairy		
RT	bacterium streptococcus thermophilus.";		
RL	Nat. Biotechnol. 22:1554-1558(2004).		
DR	EMBL; CP000023; AAV60870.1; -; Genomic_DNA.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.		
DR	GO; GO:0005506; F:iron ion binding; IEA.		
DR	GO; GO:0000287; F:magnesium ion binding; IEA.		
DR	GO; GO:0016491; F:oxidoreductase activity; IEA.		
DR	GO; GO:0006779; P:porphyrin biosynthesis; IEA.		
DR	InterPro; IPR006638; Elp3/MiaB/NifB.		
DR	InterPro; IPR010723; HemN_C.		
DR	InterPro; IPR004559; HemN_rel.		
DR	InterPro; IPR007197; Radical_SAM.		
DR	Pfam; PF06969; HemN_C; 1.		
DR	Pfam; PF04055; Radical_SAM; 1.		
DR	SMART; SM00729; Elp3; 1.		
DR	TIGRFAMs; TIGR00539; hemN_rel; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 378 AA; 43467 MW; AB25FBB520786230 CRC64;		
Query Match 13.2%; Score 262; DB 2; Length 378;			
Best Local Similarity 26.7%; Pred. No. 6.6e-10;			
Matches 100; Conservative 71; Mismatches 142; Indels 62; Gaps 17;			
Qy	12	SLYINLSPCK-----DFS---IFNRILEELKCHLI-----LLGHPPIKTLYKRVDFCLS	59
Db	7	SAYVHIPCTQICYCDFSKVPKQPVDDYLRALIREWELSDIKELRTLYIGGTPTAI	66
Qy	60	RQDNKLFITSLSKYINILEEFTLEIIPGVDFEKFPLDDEFCTIRINLVQSFSLF	119
Db	67	SAEQDLYLLSHLOKNDLSKLEEFTEIANPGDLTVDKIEVLKGSANVRVSLGVQTFDDKH	126
Qy	120	RKIVGIPISVKKLNL--INNIRKFPF--DLNIDMTVMNPKQKSHLRDLKELLSYMPE	176
Db	127	LRQIG--RSHNQAIYESIDLSKAGFNISIDLIYALPGQTMQVKNVRKALELDIP	183
Qy	177	HICFSDFICEBEGFVLRDF-----DNSIDSEKLWFCALCLESNGYINVEITNFALK	228
Query Match 13.0%; Score 258.5; DB 2; Length 379;			
Best Local Similarity 22.8%; Pred. No. 1.2e-09;			
Matches 91; Conservative 78; Mismatches 133; Indels 97; Gaps 15;			
Qy	12	SLYINLSPCK-----DFSIFNRILEELKCHLIILGHPI-----IKT	48
Db	7	SAYVHIPFCHICYCDFA-----KVLMTGQPIDAYIESLIEEFPQSFEIKLRT	55
Qy	49	LYXKVDVCLSRQDNLKFIFTSLSKYINILEEFTLEIIPGVYDFEKFLLDBFCITRI	108
Db	56	IYIGGGTPSVLSAQQLERLLTAELQDLVLEFTEANPGDLSDEVIKVLADSAVNR	115
Qy	109	NLNVQSFSLFRKIVGIPISVKKLNLINNIRKFPF--DLNIDMTVMNPKQKSHLRDL	167
Db	116	SLGVQTFNNALKKIGRTHTEVQVYD--SVRELKAGFENITIDLIYALPGQTMEMVKS	174
Qy	168	KELLSYMPHEICFSDFICEB-----SGFV--LRDFDNSIDSEKLWFCALCLESNGY	217

```
Db 175 EKFLKLPVALYSLILEDHVPMQRQRGLRLPSEDKNAD---NVEYIMDLAKNGY 231
Qy 218 INYEITPALKGHSRHNKLWELKPHGLGLYAVSLFLCNDKNNVRALIRKTSFVKA 277
Db 232 NHYEVSFGLPGPESKHNITWDNEEYVIGAGAGYL-----AGTR----- 273
Qy 278 NNHLVTFELLEDLEFFVVFHFIQGLGTIGQVSLRALRLRFEYNEKQFFQFINYCSTLSKFF 337
Db 274 -----YKULGP-----VHYLKAAPTEKRIINEVLSKSKQIEEFNL-----GLRKK- 315
Qy 338 VFDDNIMLLKGRERFKLNF-----YLVKINH---FND 367
Db 316 ---SGVLVEKFNKFCSEKLYGEQITELINQKLLYND 351
```

RESULT 10

```
Q8P177 STRP8
ID Q8P177 STRP8 PRELIMINARY; PRT; 376 AA.
AC Q8P177
DT 01-OCT-2002 (TremBrel. 22, Created)
DT 01-OCT-2002 (TremBrel. 22, Last sequence update)
DT 01-MAR-2004 (TremBrel. 26, Last annotation update)
DE Putative coproporphyrinogen III oxidase.
GN Name=hemN; OrderedLocuNames=spvM18.1022;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.U., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Rickles S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AEO10030; AAL97657.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cyttoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF04055; Radical_SAM.
DR SMART; SM00729; Elp3; I.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome.
SQ SEQUENCE 376 AA; 43093 MW; 7236F821BE82BEE5 CRC64;
```

```
Query Match 12.9%; Score 256.5; DB 2; Length 376;
Best Local Similarity 26.1%; Pred. No. 1.6e-09;
Matches 86; Conservative 63; Mismatches 148; Indels 33; Gaps 9;

Qy 12 SLYINLSFCCK-----DFS-----IPNRILEELKCHLILGHPIIKTLVIKHYD 55
Db 7 SAYVHPFCTQICYCYCDKSVFIQNPVDVAVLKALIQBFESYDI---HD-LKTLVIGGGT 62

Qy 56 FCLSRQDNLRKFTSLSKYINLELEFTELEIIPGYVDVFEKFLDDEFCITRINLVQSF 115
Db 63 PTATATKQLEYLNHLNHLERNLNLDLEETFEANPGDLTPEKIAVLQRSANVRISLGVTQ 122

Qy 116 SLEFRKIVGIPISYKKNILNINIRKPPF-DLNIDMTVNMPLQCKSHLKRDLKELLSYM 174
```

```
Db 123 NNKQLKQIGRSH-NBQIYSTIANLKTAGPHNTSIDLIYALPGQTLDOVKENAKALALD 181
Qy 175 PEHICFSDTICEBEGFVLDRF-----DNSIDSEKLFWFCALCELESNGVINYEITNFA 226
Db 182 IPHLSIYSLILEHHTVFMKORRGKLNLPTEDEAEFYEIIE-MEANGFEHYEINFT 240
Qy 227 LKGHSRHNKLWELKPHGLGLYAVSLFLCNDKNNVRALIRKTSFVKAHHVTFEL 286
Db 241 KPGFSRHNLMYNDVVEYFCGAGAGYL--NGIRYQNRVPIQHYLKAVEAGNARLNEEV 298
Qy 287 LEDLEFFVVFHFIQGLGTIGQVSLRALRLRF 316
Db 299 LRKEEMMEELFLGLRKKTKGVSIQRFQEK 328
```

RESULT 11

```
Q878L1 STRP3
ID Q878L1 STRP3 PRELIMINARY; PRT; 398 AA.
AC Q878L1
DT 01-JUN-2003 (TremBrel. 24, Created)
DT 01-JUN-2003 (TremBrel. 24, Last sequence update)
DT 01-MAR-2004 (TremBrel. 26, Last annotation update)
DE Putative coproporphyrinogen III oxidase.
GN OrderedLocuNames=SpS1180;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055 (2003).
DR EMBL; BA000034; BAC64275.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cyttoplasm; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/MiAB/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; I.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
SQ SEQUENCE 398 AA; 45841 MW; 669A0563D1206261 CRC64;
```

```
Query Match 12.9%; Score 256.5; DB 2; Length 398;
Best Local Similarity 25.8%; Pred. No. 1.7e-09;
Matches 85; Conservative 62; Mismatches 150; Indels 33; Gaps 8;

Qy 12 SLYINLSFCCK-----DFS-----IPNRILEELKCHLILGHPIIKTLVIKHYD 55
Db 29 SAYVHPFCTQICYCYCDKSVFIQNPVDVAVLKALIQBFDSY---GIRDKTLVIGGGT 84

Qy 56 FCLSRQDNLRKFTSLSKYINLELEFTELEIIPGYVDVFEKFLDDEFCITRINLVQSF 115
Db 85 PTATATKQLEYLNHLNHLERNLNLDLEETFEANPGDLTPEKIAVLQRSANVRISLGVTQ 144

Qy 116 SLEFRKIVGIPISYKKNILNINIRKPPF-DLNIDMTVNMPLQCKSHLKRDLKELLSYM 174
Db 145 NNKQLKQIGRSH-NBQIYSTIANLKTAGPHNTSIDLIYALPGQTLDOVKENAKALALD 203

Qy 175 PEHICFSDTICEBEGFVLDRF-----DNSIDSEKLFWFCALCELESNGVINYEITNFA 226
```

```

Db 204 IPHLSLSLLEHHTVFMNKKRGKLNLPTEDELAEMFEYIISE-MEANGFEHYEISNFT 262
Qy 227 LKGHESRHNKLNWELKPHLGLGLYAVSLIFCNDKNNNVRALIRKTSFVKANNHLVTFEL 286
Db 263 KPGFESRHNLMYNDVVEYFGCGAGAGYL--NGIRYQNRVPIQHLYKAVEAGNARLNEEV 320
Qy 287 LEDLEFFVYHFIQGLGTIOGVSRLRLRF 316
Db 321 LRKEEMEEELFLGLRKKTVGSIORFQEF 350

RESULT 12
ID Q8K7R5_STRP3 PRELIMINARY; PRT; 408 AA.
AC Q8K7R5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative coproporphyrinogen III oxidase.
GN Name=hemN; OrderedLocNames=SpyW3_0673;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MAGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Beres S.B., Sylva G.L., Barban K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
DR EMBL; AE014150; AAM79280.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:000506; F:iron ion binding; IEA.
DR GO; GO:000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; F:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/Miab/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR Pfam; PF06969; HemN_C; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome.
SQ SEQUENCE 408 AA; 46987 MW; C31113FBA0B1A5AD CRC64;

Query Match 12.9%; Score 256.5; DB 2; Length 408;
Best Local Similarity 25.8%; Pred. No. 1.7e-09;
Matches 85; Conservative 62; Mismatches 150; Indels 33; Gaps 8;

Qy 12 SLYNLSFCCK-----DFS-----IFNRILEELKCHLILGHPIKTLIKHVD 55
Db 39 SAYVHPFCTQCYVDFSKVFIONQPVDAVKALIQEFDSY----GIRDLKTLIYGGT 94
Qy 56 FCLSRQDNLFKFTSLSKVINLELEFTLIIIPGYVDPEFKLLDDEFCTIRINLVQSF 115
Db 95 PTAITAKQLEYLLNHLERNLNLDLEFTIEANPGDLTPEKIAVLQRSANVRISLGVTFF 154
Qy 116 SLERFKVIGIPEISYKKNILNINIRKPPF--DLNIDMTVMNPKQKSHLRDLKELLSYM 174
Db 155 NNKQLKQIGRSH-NEEQIYSTITNLKAAGFNHISIDLIYALPGQTLQDVKNVAKALD 213
Qy 175 PEHICFSDPICEEGFVLRF-----DNSIDSEKLFWCALECSNGYINYEITNFA 226
Db 214 IPHLSLSLLEHHTVFMNKKRRGKLNLPTEDELAEMFEYIISE-MEANGFEHYEISNFT 272

```

```

Qy 227 LKGHESRHNKLNWELKPHLGLGLYAVSLIFCNDKNNNVRALIRKTSFVKANNHLVTFEL 286
Db 273 KPGFESRHNLMYNDVVEYFGCGAGAGYL--NGIRYQNRVPIQHLYKAVEAGNARLNEEV 330
Qy 287 LEDLEFFVYHFIQGLGTIOGVSRLRLRF 316
Db 331 LRKEEMEEELFLGLRKKTVGSIORFQEF 360

RESULT 13
Q99ZW6_STRPY
ID Q99ZW6_STRPY PRELIMINARY; PRT; 376 AA.
AC Q99ZW6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative coproporphyrinogen III oxidase (EC 1.3.3.3).
GN Name=hemN; OrderedLocNames=SpyL040;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL; AE006549; AAK3932.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0004109; F:coproporphyrinogen oxidase activity; IEA.
DR GO; GO:000506; F:iron ion binding; IEA.
DR GO; GO:000287; F:magnesium ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006779; F:porphyrin biosynthesis; IEA.
DR InterPro; IPR006638; Elp3/Miab/NifB.
DR InterPro; IPR010723; HemN_C.
DR InterPro; IPR004559; HemN_rel.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06969; HemN_C; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00539; hemN_rel; 1.
KW Complete proteome; Oxidoreductase.
SQ SEQUENCE 376 AA; 43111 MW; 737548C68DB358E3 CRC64;

Query Match 12.9%; Score 255.5; DB 2; Length 376;
Best Local Similarity 25.8%; Pred. No. 1.8e-09;
Matches 85; Conservative 62; Mismatches 150; Indels 33; Gaps 8;

Qy 12 SLYNLSFCCK-----DFS-----IFNRILEELKCHLILGHPIKTLIKHVD 55
Db 7 SAYVHPFCTQCYVDFSKVFIONQPVDAVKALIQEFDSY----GIRDLKTLIYGGT 62
Qy 56 FCLSRQDNLFKFTSLSKVINLELEFTLIIIPGYVDPEFKLLDDEFCTIRINLVQSF 115
Db 63 PTAITAKQLEYLLNHLERNLNLDLEFTIEANPGDLTPEKIAVLQRSANVRISLGVTFF 122
Qy 116 SLERFKVIGIPEISYKKNILNINIRKPPF--DLNIDMTVMNPKQKSHLRDLKELLSYM 174
Db 123 NNKQLKQIGRSH-NEEQIYSTITNLKAAGFNHISIDLIYALPGQTLQDVKNVAKALD 181
Qy 175 PEHICFSDPICEEGFVLRF-----DNSIDSEKLFWCALECSNGYINYEITNFA 226
Db 182 IPHLSLSLLEHHTVFMNKKRRGKLNLPTEDELAEMFEYIISE-MEANGFEHYEISNFT 240
Qy 227 LKGHESRHNKLNWELKPHLGLGLYAVSLIFCNDKNNNVRALIRKTSFVKANNHLVTFEL 286

```


Qy 330 C 330
Db 356 C 356

Search completed: January 24, 2006, 19:56:48
Job time : 83.0092 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 23.2413 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-12
Perfect score: 1981
Sequence: 1 MRVDLLPLVELSLYINLSFC.....LVKINHFNDFKVKRLRP 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfilea1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	863	43.6	175	2	US-09-830-230A-521
2	763	38.5	155	2	US-09-830-230A-522
3	250.5	12.6	376	2	US-09-583-110-3497
4	248.5	12.5	409	2	US-09-107-433-4312
5	196.5	9.9	391	2	US-09-134-001C-3690
6	196.5	9.9	393	2	US-09-107-532A-6801
7	190.5	9.6	397	2	US-09-134-000C-5885
8	181	9.1	382	2	US-09-543-681A-5208
9	173	8.7	394	2	US-09-198-452A-396
10	173	8.7	402	2	US-09-438-185A-382
11	152.5	7.7	387	2	US-09-489-039A-11948
12	150.5	7.6	389	2	US-09-328-352-8088
13	143.5	7.2	506	2	US-09-352-991A-23066
14	142.5	7.2	401	2	US-09-302-540-14907
15	142	7.2	460	2	US-09-198-452A-959
16	142	7.2	460	2	US-09-438-185A-891
17	140.5	7.1	424	2	US-09-252-991A-20695
18	128	6.5	479	2	US-09-540-236-2194
19	123.5	6.2	5037	2	US-09-424-783-4
20	118.5	6.0	4968	2	US-09-424-783-5
21	118.5	6.0	5032	2	US-09-538-092-979
22	116.5	5.9	487	2	US-09-543-681A-6824
23	116.5	5.9	494	2	US-09-489-039A-12255
24	113.5	5.7	491	2	US-09-248-796A-19312
25	113.5	5.7	701	2	US-09-710-279-710
26	113.5	5.7	703	2	US-09-134-001C-3015
27	113.5	5.7	1381	2	US-09-662-254B-20

28	112.5	5.7	309	2	US-09-830-230A-121	Sequence 121, App
29	110	5.6	471	2	US-09-328-352-5145	Sequence 5145, Ap
30	109.5	5.5	871	2	US-09-792-024-81	Sequence 81, Appl
31	108.5	5.5	913	2	US-09-248-796A-15734	Sequence 15734, A
32	106	5.4	369	2	US-09-543-681A-7314	Sequence 7314, Ap
33	106	5.4	904	2	US-09-198-484-2	Sequence 2, Appli
34	105.5	5.3	287	2	US-09-830-230A-122	Sequence 122, App
35	105.5	5.3	471	1	US-08-477-451-20	Sequence 20, Appl
36	105.5	5.3	618	2	US-09-299-378-4	Sequence 4, Appli
37	104	5.2	899	2	US-09-248-796A-15467	Sequence 15467, A
38	104	5.2	2026	2	US-09-487-558B-86	Sequence 86, Appl
39	104	5.2	4866	2	US-09-424-783-2	Sequence 2, Appli
40	104	5.2	4872	2	US-09-424-783-3	Sequence 3, Appli
41	103.5	5.2	387	2	US-09-248-796A-22214	Sequence 22214, A
42	102.5	5.2	425	2	US-09-134-001C-5654	Sequence 5654, Ap
43	102.5	5.2	1226	1	US-08-540-804-12	Sequence 12, Appl
44	102.5	5.2	1226	1	US-08-218-265-12	Sequence 12, Appl
45	102.5	5.2	1226	2	US-08-521-872-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-830-230A-521
; Sequence 521, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 521
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-830-230A-521

Query Match 43.6%; Score 863; DB 2: Length 175;
Best Local Similarity 98.2%; Pred. No. 8.2e-80;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MRVDLLPLVELSLYINLSFCCKDFSIKRNILBELKCHLILGHPIIKTIYIKHVDCLSR 60
Db 1 MRVDLLPLVELSLYINLSFCCKDFSIKRNILBELKCHLILGHPIIKTIYIKHVDCLSR 60
QY 61 QNLIKFIPTSLSKYINLELLEEFTELEIIPGYVDPEFKLLDEFCITRINMLNVQSFSLR 120
Db 61 QNLIKFIPTSLSKYINLELLEEFTELEIIPGYVDPEFKLLDEFCITRINMLNVQSFSLR 120
QY 121 KIVGPIEISYKKILNINIRKPPPPDLNIDMTVMNPLQKKSHUKRDLKEL 170
Db 121 KIVGPIEISYKKILNINIRKPPPPDLNIDMTVMNPLQKKSHUKRDLQRI 170

RESULT 2
US-09-830-230A-522
; Sequence 522, Application US/09830230A
; Patent No. 6902893
; GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Lyme Disease Vaccines
FILE REFERENCE: PB481US
CURRENT APPLICATION NUMBER: US/09/830,230A
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: PCT/US98/12718
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/057,483
PRIOR FILING DATE: 1997-09-03
PRIOR APPLICATION NUMBER: 60/053,344
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/053,377
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/050,359
PRIOR FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 756
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 522
LENGTH: 155
TYPE: PR1
ORGANISM: Homo sapiens
US-09-830-230A-522

Query Match 38.5%; Score 763; DB 2; Length 155;
Best Local Similarity 98.0%; Pred. No. 1.1e-69;
Matches 147; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 21 CKDPSIFNRILEELKCHLILGHPIIKTYIKHVDPCLSRQDNLKFTTSLSKYINLELL 80
Db 1 CKDPSIFNRILEELKCHLILGHPIIKTYIKHVDPCLSRQDNLKFTTSLSKYINLELL 60

Qy 81 EEFTLEIIPGVDFEKEKLDFECITRINLNQVSFSLFEKRVIGIPISYKKNILINNI 140
Db 61 EEFTLEIIPGVDFEKEKLDFECITRINLNQVSFSLFEKRVIGIPISYKKNILINNI 120

Qy 141 RKFPFDLNDMTNMPLOKKSHLKRDLEL 170
Db 121 RKFPFDLNDMTNMPLOKKSHLKRDLELQRI 150

RESULT 3
US-09-583-110-3497
Sequence 3497, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PATH00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 3497
LENGTH: 376
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-583-110-3497

Query Match 12.6%; Score 250.5; DB 2; Length 376;
Best Local Similarity 26.0%; Pred. No. 6e-17;
Matches 89; Conservative 57; Mismatches 151; Indels 45; Gaps 10;

Qy 12 SLYINLSPCK-----DF-----STFNRIELKCHLILGHPIIKTYIKHVD 55
Db 7 SAYWHIPFCTQICYCDPSKVFIRKQPNQVDSYLEHLEEFRSYDI-----EKLSTLYIGGGT 62

Qy 56 FCLSRQDNLKFTTSLSKYINLELEEFTELEIIPGVDFEKEKLDFECITRINLNQVSF 115

Db 63 PTALSAPQLEVLNGLTKNLDLSVLEELTEANPGDLADKIAVLKNSAVNRVSLGVQTF 122
Qy 116 SLEFRKIVGIPISYKKNILINNIKKFPFD-LNIDMTNMPLOKKSHLKRDLELLSYM 174
Db 123 DDKMLKKIGRSHLE-KDIYENIDRLKLAGFDNISIDLIYALPGQWTWEQVKENYAKAIGLD 181
Qy 175 PPHICPSDFICEEGEVLRFD-----DNSIDSEKLMFCALECLECSNGYINVEITNFA 226
Db 182 IPHMSLYSILLENHTVFMNMRGKLPKPKEELEAEAFYIIAE-LERAGFEHYEISNFS 240
Qy 227 LKGHESRHNKLNWELKPHLGLGLYAVSLFCNDKNNVR-----ALIRKTGTSFYKANNHLV 282
Db 241 KPGFESRHNLMYWDNAEYVIGAGASGYV-----NGVRYKNGHPIRHYLSAVEEGNACI 294
Qy 283 TFELLEDLEFFVYHFQGLGTIGQVSLRALRLRFEYNEKQFF 324
Db 295 TEDHLSQKEQMEEMFLGRKKSGVSM-----ARFEKFGQSF 332

RESULT 4
US-09-107-433-4312
Sequence 4312, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4312:
SEQUENCE CHARACTERISTICS:
LENGTH: 409 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...409
SEQUENCE DESCRIPTION: SEQ ID NO: 4312:
US-09-107-433-4312

Query Match 12.5%; Score 248.5; DB 2; Length 409;

	Best Local Similarity	25.7%, Pred. No. 1.le-16;		
	Matches	88; Conservative	58; Mismatches	151; Indels
	Gaps	45;		
QY	12	SLYINLSFCCK-----DF-----SIFNRILBELKCHLLIGHPIIKTIYKHVD	55	
Db	40	SAYHHIPFCIQICYCDPSKVFKNPQVDSYLEHLEEFRSYDI-----EKUSTLIYRGWN	95	
QY	56	FCLSRQNKLKFIFTSLSKYINLELLEBFTLRIIPGYVDFEKFPLIDFECITRINLNQSF	115	
Db	96	TTALSAPQLDVINGLINLDLSVLSELTIEANPGDLADAKI AVLKSNVNRVSLGVTF	155	
QY	116	SLBFRKIVGIPEISYKKNLILINNIRKPPD-LNIDMTVNMPLOQKSHLRDKELLSYM	174	
Db	156	DDRWMLKKIGRSHLE-KDIYENIDRLKLAGFDNISIDLIALPGQTMEQVENVAKAIGD	214	
QY	175	PEHICFSDFICEEBGGFVLDRP-----DNSIDSEKLWFOALECSLENGVINYEITNFA	226	
Db	215	IPHMSLYSLILENHTVPFNMRMRGKLPDPKEELEAEFEYIIAE-LERAGFEHYEISNFS	273	
QY	227	LKGHESHRKNLWELKPHGLGLGLYAVSLLFCDKNKNVR-----ALIRKTGSFVKANNHLV	282	
Db	274	KPGFESRHNLMYDNAEYYIGIGAGAGYV-----NGVRYKNHGFI RHYLSAVEEGNACI	327	
QY	283	TPELLEDELFPVVYHFIOGLGTIQVSLRALRLRFYNEKQPF	324	
Db	328	TEDHLSOKEQEEMBFGLRKKSQYSVM-----ARPEETFGOSF	365	

RESULT 5

```

US-09-134-001C-3690
; Sequence 3690, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3690
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3690

```

Query Match	9.9%;	Score 196.5;	DB 2;	Length 391;
Best Local Similarity	22.9%;	Pred. No. 2e-11;		
Matches	86;	Conservative 77;	Mismatches 146;	Indels 67; Gaps 15

Qy	12	SLVINLSFCCK-----DPSIF-----	NRLEBKCHLILGHPIIKTLIKHVDCLSL 59
		: : : : : : : : : : : : :	
Db	22	SAYIHIFCYRICTYCDNFYIDKQPDVEYLDALISEMQVKNRDLRTMTVGGCTPAL 81	
Qy	60	RQDNLKKIFTFSLSKVINLELLEERTLEIIPGVDPFEKFLDEFCITITNLNVQSFSLEP 119	
		: : : : : : : : : : : : : :	
Db	82	NMRQLEKLLRAINQFTTIS--GEFSFEANPDLETYKVALLKQGVNRRISMGVOTFKPEL 139	
Qy	120	RKIVGIPESYKCLNI--LNNIRKFPDP-LNIDMTVMNPQKSHLRDLKELLSYMPE 176	
		: : : : : : : : : : : : :	
Db	140	LKILG---RTHKTEIDYDAVSHAKAGHIESLSDLMYHLPOQTDDFKDSLRALALDID 196	
Qy	177	HICFSDPICE-----EGFVLRDPDNSIDSEKLMFCALECLSESGYNIETINPAL 227	
		: : : : : : : : : : : : : :	
Db	197	HISSYGLILEPKTYNLYRKGLH--KLPNEDLGAEMYRFLMERKMSPFQHYEISLNFEGK 254	
Qy	228	KGHESRRNKUNWELKPHLGLGLYAVLSLLFCNDKNNVR-ALIRKTTGSPVKANNH----- 280	
		: : : : : : : : : : : : : :	
Db	255	LNHESNENKVVWLNREYYGFCAGASGYV-----NGVRYTNLNPVNHVYIKAINEGKCKPIL 308	

```

QY 281 ---LVTFELLEDLEFFVYHFTQGLTGTCVSLRALRLEFVYNEKQFPQFQINVCSTLSKKF 337
Db 309 SETSPTNEMREEMFEL-----GLRMNQGV-----KSRF---KKCFNKLIDE----- 348
QY 338 VFDDNIMLLKGRERFK 353
Db 349 VFGETIKDRCRGLIK 364

RESULT 6
US-09-107-532A-6801
; Sequence 6801, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUT
;
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;

```

RESULT 6

```

US-09-107-532A-6801
; Sequence 6801, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; THE INVENTION OF ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6801:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 393 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...393
; SEQUENCE DESCRIPTION: SEQ ID NO: 6801:
US-09-107-532A-6801

```

```

Query Match      9.9%; Score 196.5; DB 2; Length 393;
Best Local Similarity 21.1%; Pred. No. 2e-11;
Matches 81; Conservative 71; Mismatches 144; Indels 87; Gaps 15

Qy 2 RVDLLPLV-----ELSYINLSFC-----CKDSIF-----NRILEELKCHLL 40
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 RISMETINPAENKVSAYIHPPCEHICYCDFKNKVFLEGQPVDEYVEMLLNEMK--IML 60
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 41 LGHPI--IKTLXI-----KHVDFCLSRQDNLKPITSLSKVINLELEETLEII 88
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 ESHPTANLETILYGGGTPTSLSAKQLDRLLS-----GAREILFPQNGNEFTVEAN 110
   || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Qy 160 KSHLKRDLKELLSTWPEHICFSDF-----ICBEEGFVLRFDFNSIDS 201
Db 189 LEIFLSDLHQALUTPITHSYNLNIDTPHTGYKGRKILVPTIAQEE--ILAE--SLIA 244
Qy 202 EKLWFCALCELESNGYINYEITNFALKGHEGRHNKLINWELKPHLGLGYAVSLLFCNDKN 261
Db 245 ENL-----LLSQGFQRYELASAKPDYPKAKHNLYYTWDTRFFLGLGVASQYLH--GERS 296
Qy 262 NN-----VPALRKTGSGFVKANNHLVTFELLELEFFVYHFIOGLGTIOGVSLRALRL 314
Db 297 KNYSHISHYLRA--VRKNLPTQETSILPKGRKEIKALALRLLLEGADLAEPPSTLIGML 355
Qy 315 RFEYNEKQFFQFINYCSTLSKK--FVFDDNI 343
Db 356 TDQVKLQNLFSVHGQCLALNRQGRLFHDTI 385

RESULT 10
US-09-438-185A-382
; Sequence 382, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; TITLE OF INVENTION: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 382
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; OTHER INFORMATION: CPn0380
US-09-438-185A-382

```

Qy	315	RFEYNEKQFFQFINYCSTLSKK-FVFDNI	343
	:	:::	
Db	364	TQDVKLQNLFSVHGQCLALNRQGRLFHDTI	393
	:	:::	
RESULT 11			
US-09-489-039A-11948			
; Sequence 11948, Application US/09489039A			
; Patent No. 6610836			
; GENERAL INFORMATION:			
; APPLICANT: Gary Breton et. al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA			
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 2709.2004001			
; CURRENT APPLICATION NUMBER: US/09/489,039A			
; CURRENT FILING DATE: 2000-01-27			
; PRIOR APPLICATION NUMBER: US 60/117,747			
; PRIOR FILING DATE: 1999-01-29			
; NUMBER OF SEQ ID NOS: 14342			
; SEQ ID NO 11948			
; LENGTH: 387			
; TYPE: PRF			
; ORGANISM: Klebsiella pneumoniae			
US-09-489-039A-11948			
Query Match 7.7%; Score 152.5; DB 2; Length 387;			
Best Local Similarity 21.8%; Pred. No. 6e-07;			
Matches 7%; Conservative 57; Mismatches 143; Indels 59; Gaps 12;			
Qy	8	LVELSLYINLSFCCK-----DFS-----IFNRILEELKCHLILGHPIIKTLY	50
	:	:::	
Db	12	LPPLSLYIHIPWCVKQKPCYDFNSHALKGEVPHDDYVQHLLNDLQADAHYAQGREGITGF	71
	:	:::	
Qy	51	IKHVDFCLSRQDNLKFTFTSLSKYINLELBEFTLEIPGYVDFPEKFLDDEFCTIRNL	110
	:	:::	
Db	72	IGGTPSLLSGPAMQTLIDGVRACLPAAGAEITMEANPGTVEADRFVDYQAGVNRISI	131
	:	:::	
Qy	111	NVQSFS-----LEFRKIVGIEISYKKNLINLNIRKFPFDLNDMTVMNPLQKSHLKR	165
	:	:::	
Db	132	GVQSFSFKQLRGRING-PPEAKRAARLASGLRSF-----NLDLMHGLPDQSLERALD	186
	:	:::	
Qy	166	DLKELLSYMPHEHICFSDFICEEGFVLDRFDNSI-----DSEKLFWC-----ALSCLE	213
	:	:::	
Db	187	DLRQAIANPPLHSWYQLTIE-----PNTLFGSRPPVLPDDDLNDLWDFEGCHQLLS	237
	:	:::	
Qy	214	SNGYINTYEITNFALKGHESRHNKLNWELKPHLGLGYAVSLIFCND-----KNNNVRA	266
	:	:::	
Db	238	AAGVQVQVETSAKAPGPGQCQHNLYWFGDYLIGCGAGHKITFPDGRILRTAKTRHPRG	297
	:	:::	
Qy	267	LIRKTSGFVKANNHLVTFELLEDLFFVYVHF	297
	:	:::	
Db	298	YME--GRYLEROHQDVE--EADKPPFFPMNRF	324
	:	:::	

Search completed: January 24, 2006, 20:02:42
Job time : 24.2413 secs

• This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 66.2599 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-12
Perfect score: 1981
Sequence: 1 MRVDLLPLVLSLYINLSFC.....LVKINHFNDFPKVRLRP 377

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues
Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
1	1981	100.0	377 5	US-10-688-058-12 Sequence 12, Appl
2	1090	55.0	204 5	US-10-688-058-10 Sequence 10, Appl
3	863	43.6	175 5	US-10-994-726-521 Sequence 521, Appl
4	763	38.5	155 5	US-10-994-726-522 Sequence 522, Appl
5	250.5	12.6	376 5	US-10-472-928-2860 Sequence 2860, Ap
6	248.5	12.5	409 5	US-10-617-320-4312 Sequence 4312, Ap
7	231.5	11.7	385 5	US-10-470-048B-106 Sequence 106, Appl
8	206.5	10.4	381 5	US-10-650-274-98 Sequence 98, Appl
9	173	8.7	394 4	US-10-289-762-396 Sequence 396, Appl
10	168	8.5	352 4	US-10-335-977-6718 Sequence 6718, Ap
11	168	8.5	364 4	US-10-335-977-6719 Sequence 6719, Ap
12	165	8.3	345 4	US-10-335-977-6717 Sequence 6717, Ap
13	150.5	7.6	493 4	US-10-425-115-261185 Sequence 261185, Ap
14	142	7.2	460 4	US-10-289-762-959 Sequence 959, Appl
15	135.5	6.8	819 4	US-10-282-122A-47000 Sequence 47000, A
16	129.5	6.5	1034 5	US-10-732-923-23287 Sequence 23287, A
17	128	6.5	1219 5	US-10-732-923-23209 Sequence 23209, A
18	124.5	6.3	548 4	US-10-724-972A-7488 Sequence 7488, Ap
19	124.5	6.3	605 4	US-10-282-122A-54636 Sequence 54636, A
20	123	6.2	665 4	US-10-437-963-172583 Sequence 172583, A
21	123	6.2	701 4	US-10-282-122A-47140 Sequence 47140, A
22	121	6.05	4	US-10-238-075-691 Sequence 691, Appl
23	120.5	6.1	2055 4	US-10-276-774-1795 Sequence 1795, Ap
24	118.5	6.0	4767 4	US-10-276-774-1902 Sequence 1902, Ap
25	118.5	6.0	4967 4	US-10-668-767-60 Sequence 60, Appl
26	118.5	6.0	4967 4	US-10-668-767-62 Sequence 62, Appl
27	118.5	6.0	4968 4	US-10-668-767-61 Sequence 61, Appl

28	118.5	6.0	5032 4	US-10-408-765A-26 Sequence 26, Appl
29	118.5	6.0	5038 5	US-10-723-860-1310 Sequence 1310, Ap
30	118.5	6.0	5038 5	US-10-756-149-5039 Sequence 5039, Ap
31	118.5	6.0	5081 4	US-10-276-774-1850 Sequence 1850, Ap
32	118.5	6.0	5081 5	US-10-450-763-35328 Sequence 35328, A
33	113.5	5.7	703 4	US-10-724-972A-3982 Sequence 3982, Ap
34	113.5	5.7	1381 5	US-10-706-635-20 Sequence 20, Appl
35	112.5	5.7	309 5	US-10-994-726-121 Sequence 121, Appl
36	112	5.7	1478 5	US-10-732-923-3353 Sequence 3353, Ap
37	111.5	5.6	742 4	US-10-437-963-109290 Sequence 109290, A
38	111.5	5.6	950 4	US-10-282-122A-47285 Sequence 47285, A
39	111.5	5.6	2326 5	US-10-732-923-22500 Sequence 22500, A
40	110	5.6	1935 5	US-10-732-923-13992 Sequence 13992, A
41	109.5	5.5	871 4	US-10-032-585-7050 Sequence 7050, Ap
42	109.5	5.5	871 5	US-10-882-104-81 Sequence 81, Appl
43	109	5.5	485 4	US-10-282-122A-52735 Sequence 52735, A
44	109	5.5	933 4	US-10-282-122A-47180 Sequence 47180, A
45	108.5	5.5	1032 4	US-10-607-631-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-688-058-12
; Sequence 12, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHERINE F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; BORRELIA
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:00305
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; TYPE: PRT
; LENGTH: 377
; ORGANISM: Borrelia burgdorferi
US-10-688-058-12

Query Match				100.0%;	Score 1981;	DB 5;	Length 377;
Best Local Similarity				100.0%;	Pred. No. 8.5e-169;		
Matches 377;				Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MRVDLLPLVLSLYINLSFCCKDFSTFNRILBELKCHLILGHPIIKTLVKKHVDVFCLSR	60				
Db	1	MRVDLLPLVLSLYINLSFCCKDFSTFNRILBELKCHLILGHPIIKTLVKKHVDVFCLSR	60				
Qy	61	QDNLKFTISLSKYINLELEEFFLEIIPGVDPPEFKLDFECITRINLVQSFSLEFR	120				
Db	61	QDNLKFTISLSKYINLELEEFFLEIIPGVDPPEFKLDFECITRINLVQSFSLEFR	120				
Qy	121	KIVGPIEISYKGLNLIINNIRKFPFDLIDMTVMNPLQKSHLKRDLKELLSYMPHEICF	180				
Db	121	KIVGPIEISYKGLNLIINNIRKFPFDLIDMTVMNPLQKSHLKRDLKELLSYMPHEICF	180				
Qy	181	SDFCIEBEGFVLDFDONSIDSEKLFWFCALCSLNGYINTEITNFALKGHESHKLNWE	240				
Db	181	SDFCIEBEGFVLDFDONSIDSEKLFWFCALCSLNGYINTEITNFALKGHESHKLNWE	240				
Qy	241	LKPHLGLYAVSLFCNDKNNVRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIOG	300				
Db	241	LKPHLGLYAVSLFCNDKNNVRALIRKTSFVKANNHLVTFELLEDLEFFVYHFIOG	300				
Qy	301	LGTTQGVSLRALRURPEYNEKQFQFINYCSTLSKFKVFDDNIMLLKGRERFKLNFYLK	360				
Db	301	LGTTQGVSLRALRURPEYNEKQFQFINYCSTLSKFKVFDDNIMLLKGRERFKLNFYLK	360				

```
Qy 361 IINHFNDFKVKLRLP 377
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
Db 361 IINHFNDFKVKLRLP 377
US-10-994-726-521

Query Match 43.6%; Score 863; DB 5; Length 175;
Best Local Similarity 98.2%; Pred. No. 4.4e-69;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRVDLLPLVELSLYINLSFCCCKDPSIFNRIIEELKCHLILGHPIIKTYIKHVDVFCLSR 60
Db 1 MRVDLLPLVELSLYINLSFCCCKDPSIFNRIIEELKCHLILGHPIIKTYIKHVDVFCLSR 60
Qy 61 QDNLKFIPTSLSKVINLELEBFTLEIIPGYVDFPEFKLDEFCITRINLNVSQFSLEFR 120
Db 61 QDNLKFIPTSLSKVINLELEBFTLEIIPGYVDFPEFKLDEFCITRINLNVSQFSLEFR 120
Qy 121 KIVGIPEISYKLNILNINIRKFPFDLNDMTVMPLQKSHLKRDLKEL 170
Db 121 KIVGIPEISYKLNILNINIRKFPFDLNDMTVMPLQKSHLKRDLQRI 170

RESULT 4
US-10-994-726-522
; Sequence 522, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 522
; LENGTH: 155
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-994-726-522

Query Match 38.5%; Score 763; DB 5; Length 155;
Best Local Similarity 98.0%; Pred. No. 3.4e-60;
Matches 147; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 21 CKDPSIFNRIIEELKCHLILGHPIIKTYIKHVDVFCLSRODNLKFIPTSLSKVINLELL 80
Db 1 CKDPSIFNRIIEELKCHLILGHPIIKTYIKHVDVFCLSRODNLKFIPTSLSKVINLELL 60
Qy 81 BEFTLEIIPGYVDFPEFKLDEFCITRINLNVSQFSLEFRKIVGIPEISYKLNILNINI 140
Db 61 BEFTLEIIPGYVDFPEFKLDEFCITRINLNVSQFSLEFRKIVGIPEISYKLNILNINI 120
Qy 141 RKFPFDLNDMTVMNPLQKSHLKRDLKEL 170
Db 121 RKFPFDLNDMTVMNPLQKSHLKRDLQRI 150

RESULT 5
US-10-472-928-2860
; Sequence 2860, Application US/10472928
; Publication No. US20050020813A1
```

```
Qy 361 IINHFNDFKVKLRLP 377
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
Db 361 IINHFNDFKVKLRLP 377
US-10-688-058-10

Query Match 55.0%; Score 1090; DB 5; Length 204;
Best Local Similarity 100.0%; Pred. No. 2.5e-89;
Matches 204; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 RKFPFDLNDMTVMNPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 200
Db 1 RKFPFDLNDMTVMNPLQKSHLKRDLKELLSYMPHEICFSDFCIEEGFVLRDFDINSID 60
Qy 201 SEKLWFCALCELENGYINYEITNFALKGHESRNKNLWELKPHLGLGLYAVSLFCNDK 260
Db 61 SEKLWFCALCELENGYINYEITNFALKGHESRNKNLWELKPHLGLGLYAVSLFCNDK 120
Qy 261 NNNVRALIRKTSFVKANNHLVTFELLEDLFFVYHFTQGLGTIQGVSLRALRFRFYNE 320
Db 121 NNNVRALIRKTSFVKANNHLVTFELLEDLFFVYHFTQGLGTIQGVSLRALRFRFYNE 180
Qy 321 KQFFQFNYCSTLKKFVDDNIM 344
Db 181 KQFFQFNYCSTLKKFVDDNIM 204

RESULT 3
US-10-994-726-521
; Sequence 521, Application US/10994726
; Publication No. US20050147999A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481D1
; CURRENT APPLICATION NUMBER: US/10/994,726
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: 09/830,230
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 521
```



```

; SEQ ID NO 106
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-106

Query Match          11.7%; Score 231.5; DB 5; Length 385;
Best Local Similarity 24.3%; Pred. No. 4.7e-12;
Matches 94; Conservative 79; Mismatches 153; Indels 61; Gaps 16;

Qy 12 SLVINLSFCCK-----DFSIP-----NRILEELKCHLILLGHPIKTKLYIKHVDFCLS 59
Db 16 SAYIHIFPCVRICTYCDNFYFIQNPQVDEYLDALITEMSTAKYRILKTMVVGSTPTAL 75

Qy 60 RQDNLFKIFTLSKYINLELLEEFTRIIPGYVDFEKFPLDEFCTRINLNVOFSLSLF 119
Db 76 SINGLERLLKAIRD--FTTITGEYTFEANPDELTKVKQLLEKYGVKRISGVQTFKPEL 133

Qy 120 RKIVG-----IPETSYKKNLINIRKPPFDLNDMTVMNPLQKKSHLKRDLKELLSTMP 175
Db 134 LSVLGRTHNTEDIYTSVLNKNAGIK---SISLDLMYHLPKQTIPEQSLDLALDMDI 189

Qy 176 EHICFSDFICE-----EEGFV--LEDPDNSIDSEKLWFCALCESNGYINVEITNF 225
Db 190 QHISSYGLIIEPKTQFYVMYRKGLUKLPNEDLGADMYQL---LMSKIEQSPFHQTEISNF 246

Qy 226 ALKGHESHRNKNLWELPHGLGLGYA-----VSLLCNDKNNNVRALIRKTGSPVKANN 279
Db 247 ALDGHESEHKVYWFNEEYGFAGAGSYVDGVRYTNPNVNYHIKAINKESKAILVSNK 306

Qy 280 HLVTFFELLEDLPFVYVHPIQGLGTIQGVSLRALRLRFEYN-EKOFFQPINYCSTLSKKFV 338
Db 307 PSLT-ERMEEB-EMEL-----GLRLNEGVSRRFKKFKQSTIESVFGQTIN--NLKEKELI 357

Qy 339 FDDN-IMLXGR-----ERFKLN 355
Db 358 VERKNDVIALTRNGKVGIGNEVFEAFLLN 384

RESULT 8
US-10-650-274-98
; Sequence 98, Application US/10650274
; Publication No. US20050202437A1
; GENERAL INFORMATION:
; APPLICANT: GLENN, MATTHEW
; APPLICANT: HAVUKKALA, ILKKA J
; APPLICANT: LUBBERS, MARK WILLIAM
; APPLICANT: DEKKER, JAMES
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES,
; TITLE OF INVENTION: MATERIALS INCORPORATING THEM, AND METHODS FOR USING
; TITLE OF INVENTION: THEM.
; FILE REFERENCE: 11000.1073
; CURRENT APPLICATION NUMBER: US/10/650, 274
; CURRENT FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0
; SEQ ID NO 98
; LENGTH: 381
; TYPE: PRT
; ORGANISM: LACTOBACILLUS RHAMNOSUS
US-10-650-274-98

```

	Query Match	10.4%	Score 206.5	DB 5	Length 381
	Best Local Similarity	24.7%	Pred. NO. 8e-10		
	Matches 89	Conservative 63	Mismatches 121	Indels 87	Gaps 17
Qy	14 YINLSPC---CKQFSIFNRI-----LEELKCHLILGHP--IKITLYI---51				
Db	5 YIHIPFCEHICY-YCDFKNKVFIEGQVDVYVAMLLKEMR--MVMAHEPPEEKIETVYVGGG61				
Qy	52 ---KHVDFCLSRDNDLKFTITSLSKVINLELLEBEFTLEIIPG-VYDFEPEKLDE102				
Db	62 TPPTTLTPQOLAVICGIGRIILHEDH-----GEFTFANPNDLITTKLOVLVD109				

```

Qy 103 FCITRNLNVQSPSLFRKIVGIPETISYKKNLTLNNIRKFPED-LNIDMTVANPQKKS 161
Db 110 FGVNRLSIGVQSFNDVLKRIGRHRA-KDVTYTAIGNARKVGFNDLSIDLIFRLPDQSRD 168
Qy 162 HLKRDLEKLLSYMPEHICFSDFICEBEGFVLDRFDNSIDSEKL-----WFCAL 210
Db 169 DFLNSIQKALADLP HYSYTSILERKTI-----FYNLMRQRKLRLPTQDV EADMYQDAID 224
Qy 211 CLBSNGYVINEITNFALKGHESRHNKLINWELKPHLGLGYAVSLLFCNDRKNNVRLIRK 270
Db 225 LMEAHGRHQYEISNFATKYQCRHNLLYWQNDKYFGGAGAGYL-GRDRYHNYGPIKQY 283
Qy 271 TGSF-----VKANNHLVPELLEDLEFFVYHFIQGLGTIQGYSLARLRFEYNEKQFF 324
Db 284 LAPLHADHLPVLAHLVVPVSEQIEE-EMFL-----GLRTWAGV-----NEDRFY 326

RESULT 9
US-10-289-762-396
; Sequence 396, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffois, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides.
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 396
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; US-10-289-762-396

```

```

Query Match      8.7%; Score 173; DB 4; Length 394;
Best Local Similarity 23.1%; Pred. No. 8.4e-07;
Matches 90; Conservative 68; Mismatches 164; Indels 68; Gaps 16;

Qy      3 VDLPLV-----ELSLYNLSFC-----CKDPSI-FNRILEEKCHLILG-----HPI 45
Db      15 VSLPLMNGKAPLALYIHIFPTCKCRYCSFTIIPKSESVSLYNAVIOEGRKLAPI 74

Qy      46 KSLYIKHVPD-----CLSRQDNLKIFTSLSKYINLELEEEFTLEIIPGVDFEKPFL 100
Db      75 QETHFIEVFFGGTSLVSPGLDKRLIKELAPH-----AREITLEANPENLTVSYLRQL 129

Qy      101 DECITRINLNVQSFSLPEFKIYGVIPSEIYKKNLILNNIRKPPF-DLNDTMTVNMPLQK 159
Db      130 QETPINRISGVGTDFDSDIILQILGRTHSSAAITAL-QEONHGFSNLSIDLIYGLPTQS 188

Qy      160 KSHLRDLKELLSVMPHEHICFSDP-----IICEEGFVLRLDFDSDIDS 201
Db      189 LEIFSLDLHQALTLPITHISLYNLTDPHTSFYKHKILVPTIAQEE--ILAEM--SLLA 244

Qy      202 EKLWFCALCELENGYINYEITPALKGHESRHKNLWELKPHIGLGYAVSLFCNDKN 261
Db      245 ENL-----LLSQGFORYELASYAKDPYPAKHNLYYVTDTPFIFGLGVSAQYLH-GERS 296

Qy      262 NN-----VRALIRKTGSFVKANNHVLVTFELLEDELFVVFVHFQGLGTTQGVSLRALRL 314
Db      297 KNYSHISHYLRA-VKNNLPQETSEIILPKKERIKEALALRLLEGADLAEPFTLSIML 355

Qy      315 RFEYNEKQFTQFINYCSTLSKK-FVFPDNI 343
Db      356 TQDVKLNQLFSVHGOCIALNRQGRLFHDTI 385

```

RESULT 10
US-10-335-977-6718
: Sequence 6718, Application US/10335977

Db	283	LKRQVETLNKQDKLEKFLGURCELGVELSLDENKVKFLIENKAFIK-NNRLI	337
<p>Publication No. US20040052799A1</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: DOUGLAS SMITH et al</p> <p>TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS</p> <p>NUMBER OF SEQUENCES: 10031</p> <p>CORRESPONDENCE ADDRESS:</p> <p>ADDRESSEE: LAHIVE & COCKFIELD</p> <p>STREET: 28 State Street</p> <p>CITY: Boston</p> <p>STATE: Massachusetts</p> <p>COUNTRY: USA</p> <p>ZIP: 02109-1875</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: CD-ROM ISO9660</p> <p>COMPUTER: IBM PC Compatible</p> <p>OPERATING SYSTEM: Windows NT 4.0</p> <p>SOFTWARE: UNIX</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/10/335,977</p> <p>FILING DATE: 30-Dec-2002</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: 08/993,002</p> <p>FILING DATE: 17-DEC-1997</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Mandragouras, Amy E.</p> <p>REGISTRATION NUMBER: 36,207</p> <p>REFERENCE/DOCKET NUMBER: GTN-018</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (617)227-7400</p> <p>TELEFAX: (617)742-4214</p> <p>INFORMATION FOR SEQ ID NO: 6718:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 352 amino acids</p> <p>TYPE: amino acid</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: protein</p> <p>HYPOTHETICAL: YES</p> <p>ORIGINAL SOURCE:</p> <p>ORGANISM: Helicobacter pylori</p> <p>FEATURE:</p> <p>NAME/KEY: misc feature</p> <p>LOCATION: (B) LOCATION 1...352</p> <p>SEQUENCE DESCRIPTION: SEQ ID NO: 6718:</p>			
US-10-335-977-6718			
<p>Query Match 8.5%; Score 168; DB 4; Length 352;</p> <p>Best Local Similarity 23.0%; Pred. No. 2e-06;</p> <p>Matches 82; Conservative 55; Mismatches 113; Indels 106; Gaps 15;</p>			
Qy	9	VELSLYINISFC-----CKDPSIFNR-----ILEELKCHL-----ILLGH	43
Db	6	INMILYIHIPFCNKCGYCAFNSENKHLKEEYIQAALCLDLKHALSQTDRPIESIFIG	65
Qy	44	PIIKTYIKHVDCLSRQNLKFIPTSLSKYINLELBEFTLEIIPGVYDFEKFLLDEF	103
Db	66	GTPTNLSVK-----AFERIPESIHQHASLSDMCDEITTEANPELISKAWCQGLDL	115
Qy	104	CITRINLVQSRSLERFKIVGIPETISYKKNILINNIRKFPDP-LNIDMTVNMPLQKSH	162
Db	116	GINRLSLGVQSFRED--KLLFLERQHSKNIAPIVBIIFKSGIENISIDLIYNTPLDNETS	173
Qy	163	LKRDLEKLLSYMP-EHICFSDPFICEEGFVLK-----DPDNSIDSEKLWFCALCGL	213
Db	174	LKEELK-LAKELPINHLISAYALSIEKNTNLEKNKPPSSVDFDNNV-----REVLE	223
Qy	214	SGNYINYEITNFALKGHBSRHNKAWELPHLGLG-----LYAVSL-----	255
Db	224	GFFFKQYEVSNYA-RNYQVKNLAYWGAKDYLGCGAGVGVANERFFAKKLIENYIKDP	282
Qy	256	FCNDK-----NNVRALIRKTGSPVKANNHLV	282

Qy 214 SNGVYIETNFKLGHESRHNKLNWELKPHLG- - - - -LYAVSLL- - - - - 255
Db 236 GFFPKQYEVSNYA - RNYQVKNLAYWGAKDYLGCGAGVGVANERFYAKKLIENYIKDP 294
Qy 256 - - - - -FCNDK- - - - -NNVRALIRKTSFVKNHVLV 282
Db 295 LKROVETLNKQDKCLEKFLGLRCELGVLSLLDENKVKFLIENKAFIK-NNRLI 349

RESULT 12
US-10-335-977-6717
; Sequence 6717, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 6717:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 345 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...345
; SEQUENCE DESCRIPTION: SEQ ID NO: 6717:
US-10-335-977-6717

Query Match 8.3%; Score 165; DB 4; Length 345;
Best Local Similarity 23.3%; Pred. No. 3.6e-06;
Matches 82; Conservative 53; Mismatches 11; Indels 106; Gaps 15;

Qy 13 LYINLSFC- - - - -CKDPSIFNR- - - - -LLEELKCHL- - - - -ILGHPIIK 47
Db 3 LYIHIPFCNKGCGYAFNYSYENKGLKEVYIQAICLDLKHALSQTDEPIESIFIGGGTPN 62
Qy 48 TLYIKHVDCLSRQDNLFKFTSLSKYINLELLEFTEIIPGYVDPEFKLLDDEFCTIR 107
Db 63 TLSVK- - - - -AFRIEISIHQHASLSDMCETITEANPELISKAWCOGLKDLGNR 112
Qy 108 INLVNQSFSLFRKTVIGTPEISYKKLNILNIRKFPFD-LNIDMTVMNPLQKSHLRD 166

Db 113 LSLGVQSFRED- - - - -KLLFLERQHSKNIAPIVIEIIFKSGIENISIDLIYNTPLDNETSUKBE 170
Qy 167 LKELLSYMP-EHICFSDFICEERGFVLR- - - - -DFDMSIDSEKLMFCALECSNGY 217
Db 171 LK-LAKELPINHLISAYALSIEKNTNLEKNAKSPSSVDFDNV- - - - -REVLEGGFF 220
Qy 218 INYEITNFKLGHESRHNKLNWELKPHLG- - - - -LYAVSLL- - - - - 255
Db 221 KQYEVSNYA - RNYQVKNLAYWGAKDYLGCGAGVGVANERFYAKKLIENYIKDPLKQ 279
Qy 256 - - - - -FCNDK- - - - -NNVRALIRKTSFVKNHVLV 282
Db 280 VETLNKQDKCLEKFLGLRCELGVLSLLDENKVKFLIENKAFIK-NNRLI 330

RESULT 13
US-10-425-115-261185
; Sequence 261185, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 261185
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_169815C.1.pep
US-10-425-115-261185

Query Match 7.6%; Score 150.5; DB 4; Length 493;
Best Local Similarity 21.5%; Pred. No. 0.00012;
Matches 67; Conservative 54; Mismatches 121; Indels 69; Gaps 11;

Qy 5 LLPVLVELSLYINLSFCCK- - - - -DFSIF- - - - -NRILEELKC 36
Db 67 LPPLPPASAVVHLPPFCFKCHYCDFFIALLSSSAPSRSGEAADPRIVDVYVNLLEVSVA 126
Qy 37 HLILLGHI- - - - -IKTYIKHVDCLSRQDNLFKFTSLSKYINLELLEFTEIIPG 90
Db 127 - - - - -TRPVSDGDGVPLETVFVGGTPLSVPPRLVATVLDALRARFGLSARPEVSIEMDPG 181
Qy 91 YVDPEKFLDDEFCTIRINLVQSFLEPRKIV- - - - -GIPEISYKKLNIL- - - - -INNIR 141
Db 182 TFDAAARLKLGVGVNRVSLGVQAFQEDLLRACGRANGREV-HEAVAIVTACGLQN- - - 238
Qy 142 KFPFDLNDMTVMNPLQKSHLRDLKELLSYMEPHICFSDFICEE- - - - -EGFVLDFD 196
Db 239 - - - - -WSMDLISSLPNQTEEMWESLACTVDARTHVSVYDLOIEQGTQKQMTPTGVP 293
Qy 197 NSIDSEKLF- - - - -CALECSNGYINYEITNFKLGHESRHNKLNWELKPHLGILGYAVSL 254
Db 294 LPSDTESANFVRIASKRLSEAGYNHYEISSYCKPGYECCKNLTYWQNRPFVAFGLGSASY 353
Qy 255 LFCNDKNNVR 265
Db 354 I- - - - -NGVR 358

RESULT 14
US-10-289-762-959
; Sequence 959, Application US/10289762
; Publication No. US2004006218A1
; GENERAL INFORMATION:
; APPLICANT: Grifffais, R.

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 38.8844 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-14
Perfect score: 875
Sequence: 1 QIQTHILASILESIIFILR.....IADTTVHRIIEDGVYLIRTE 174

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	875	100.0	174	8	Adu98755 Borrelia
2	872	99.7	1277	8	Adu98757 Borrelia
3	109	12.5	1629	4	Aau48711 Propionib
4	109	12.5	1629	6	Abm45230 Propionib
5	104.5	11.9	337	7	Adc01324 Enteroha
6	102	11.7	901	9	Adw16234 Mmel endo
7	102	11.7	919	8	Adi61811 Mmel rest
8	99	11.3	1032	6	Abu25689 Protein e
9	97	11.1	374	6	Aae30333 Human pro
10	97	11.1	389	8	Adl23826 Pyrococcu
11	97	11.1	1005	8	Adq18134 Human sof
12	97	11.1	1113	7	Adn39946 Cancer/an
13	97	11.1	1337	6	Abu56496 Lung canc
14	97	11.1	1337	6	Abu56690 Lung canc
15	97	11.1	1337	7	Adn39060 Cancer/an
16	97	11.1	1359	7	Adn39948 Cancer/an
17	97	11.1	1448	6	Abu56689 Lung canc
18	97	11.1	1448	6	Abu56495 Lung canc
19	97	11.1	1448	7	Adn39058 Cancer/an
20	97	11.1	1448	7	Adn39941 Cancer/an
21	97	11.1	1456	6	Abu56688 Lung canc
22	97	11.1	1456	6	Abu56494 Lung canc
23	97	11.1	1456	7	Adn39943 Cancer/an
24	97	11.1	1456	7	Adn39056 Cancer/an

25	97	11.1	1493	7	Adn39947	Adn39947	Cancer/an
26	97	11.1	1494	6	Abu56691	Abu56691	Lung canc
27	97	11.1	1494	6	Abu56497	Abu56497	Lung canc
28	97	11.1	1494	7	Adn39062	Adn39062	Cancer/an
29	97	11.1	1494	7	Adn39945	Adn39945	Cancer/an
30	97	11.1	2308	2	AAR57902	AAR57902	Human RPT
31	97	11.1	2308	4	AAB59586	AAB59586	Human rec
32	97	11.1	2308	6	ABG72433	ABG72433	Receptor
33	97	11.1	2314	6	ABR58521	ABR58521	Human pro
34	97	11.1	2314	6	ABU56692	ABU56692	Lung canc
35	97	11.1	2314	6	AAE30332	AAE30332	Human PTP
36	97	11.1	2314	7	ADBE2111	ADBE2111	Human pro
37	97	11.1	2314	7	ADE62042	ADE62042	Human pro
38	97	11.1	2314	7	ADE62038	ADE62038	Human pro
39	97	11.1	2314	7	ADN39064	ADN39064	Cancer/an
40	97	11.1	2314	8	ADU75597	ADU75597	Marker ge
41	97	11.1	2314	8	ADU06410	ADU06410	Novel bro
42	97	11.1	2314	9	ADV70177	ADV70177	Tumor-ass
43	97	11.1	2314	9	AEA00158	AEA00158	Human TAT
44	97	11.1	2314	9	AEA04969	AEA04969	Protein t
45	97	11.1	2314	9	AEA00678	AEA00678	Human TAT

ALIGNMENTS

RESULT 1
ADU98755
ID ADU98755 standard; protein; 174 AA.
XX
AC ADU98755;
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 14.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
PN W02004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
DR WPI; 2004-834155/82.
XX
DR N-PSDB; ADU98754.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises
PT providing to the subject at least one Borrelia antigen or its fragment.
XX
PS Claim 27; SEQ ID NO 14; 121pp; English.
XX
CC The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide
CC encoding a polypeptide for an ability to produce an immune response;
CC preparing a vaccine; vaccinating a subject; treating a subject infected

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 174 AA;

Query Match 100.0%; Score 875; DB 8; Length 174;
 Best Local Similarity 100.0%; Pred. No. 5.3e-79;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QITQHILASILEESIIFLIRIFFTAYIEDNDIFPKILQENKLYESSISFRFFYDENTKK 60
 DB 1 QITQHILASILEESIIFLIRIFFTAYIEDNDIFPKILQENKLYESSISFRFFYDENTKK 60
 QY 61 KLEYKIIITIFNLLDKGSDAIKFPVFNGLFSEDKVKYLNNEGLLSISEIEILVKMLFF 120
 DB 61 KLEYKIIITIFNLLDKGSDAIKFPVFNGLFSEDKVKYLNNEGLLSISEIEILVKMLFF 120
 QY 121 EEKNIKDEKFKVYKSRLDPKSGELYETLLEVDLRIADTTVHRIEDGVYLIRTE 174
 DB 121 EEKNIKDEKFKVYKSRLDPKSGELYETLLEVDLRIADTTVHRIEDGVYLIRTE 174

RESULT 2
 ADU98757
 ID ADU98757 standard; protein; 1277 AA.
 AC ADU98757;
 DT 24-FEB-2005 (first entry)
 DE Borrelia burgdorferi antigenic polypeptide segid 16.
 KW antibacterial; vaccine; immune stimulation; immunity; antigen;
 KW DNA library.
 OS Borrelia burgdorferi.
 XX WO2004103269-A2.
 XX 02-DEC-2004.
 XX 17-OCT-2003; 2003WO-US033056.
 XX 18-OCT-2002; 2002US-0419401P.
 XX (MACR-) MACROGENICS INC.
 XX (TEXA) UNIV TEXAS SYSTEM.
 XX Sykes KF, Hale KS, Johnston SA;
 XX WPI; 2004-834155/82.
 XX N-PSDB; ADU98756.
 XX Immunizing a subject against Borrelia burgdorferi infection comprises
 XX providing to the subject at least one Borrelia antigen or its fragment.
 XX Claim 27; SEQ ID NO 16; 121pp; English.

XX The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 1277 AA;

Query Match 99.7%; Score 872; DB 8; Length 1277;
 Best Local Similarity 99.4%; Pred. No. 1.5e-77;
 Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QITQHILASILEESIIFLIRIFFTAYIEDNDIFPKILQENKLYRSSISFRFFYDENTKK 60
 DB 268 EITQHILASILEESIIFLIRIFFTAYIEDNDIFPKILQENKLYRSSISFRFFYDENTKK 327
 QY 61 KLEYKIIITIFNLLDKGSDAIKFPVFNGLFSEDKVKYLNNEGLLSISEIEILVKMLFF 120
 DB 328 KLEYKIIITIFNLLDKGSDAIKFPVFNGLFSEDKVKYLNNEGLLSISEIEILVKMLFF 387
 QY 121 EEKNIKDEKFKVYKSRLDPKSGELYETLLEVDLRIADTTVHRIEDGVYLIRTE 174
 DB 388 EEKNIKDEKFKVYKSRLDPKSGELYETLLEVDLRIADTTVHRIEDGVYLIRTE 441

RESULT 3
 AAU48711
 ID AAU48711 standard; protein; 1629 AA.
 AC AAU48711;
 DT 27-FEB-2002 (first entry)
 DE Propionibacterium acnes immunogenic protein #9607.
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX Propionibacterium acnes.
 OS WO200181581-A2.
 XX 01-NOV-2001.
 XX 20-APR-2001; 2001WO-US012865.
 XX 21-APR-2000; 2000US-0199047P.
 XX 02-JUN-2000; 2000US-0208841P.
 XX 07-JUL-2000; 2000US-0216747P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI; 2001-616774/71.
 XX N-PSDB; AAS59543.
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 XX vaccinating against and diagnosing infections, especially useful for
 XX treating acne vulgaris.
 XX Example 1; SEQ ID NO 9906; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central


```

XX 24-JAN-2001; 2001JP-00112010.
XX (UYTS-) UNIV TSUKUBA.
XX WPI; 2003-451640/43.
XX Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
XX and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX Claim 3; SEQ ID NO 1369; 2067pp; Japanese.
XX The invention relates to a novel enterohaemorrhagic Escherichia coli
XX O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
XX has anti-bacterial activity. The polypeptide can be used in detection
XX and/or treatment of O157:H7 infection. The nucleotide sequence of the
XX genome of Enterohaemorrhagic E coli O157:H7 was determined. The present
XX sequence represents an E. coli O157:H7-specific polypeptide of the
XX invention.
XX Sequence 337 AA;
XX
XX Query Match 11.9%; Score 104.5; DB 7; Length 337;
XX Best Local Similarity 21.0%; Pred. No. 0.094;
XX Matches 45; Conservative 46; Mismatches 74; Indels 49; Gaps 8;
XX
Qy 1 QITOHILASILESIIFL-----RIFFIAYIEDNDIFKKILOEN-----KLYRSSISF 49
Db 57 KFTPHIADYLDKBYIELLSQATKYQTVIKLYHIDSEPLKALPQSNWPVSIYRLLSF 116
Qy 50 RYF-----FYDENTKKK-----LEYKK-----IITFNLLDKGSDAIFPVFNG 88
Db 117 DYFARLDLSLYLDADIVCKGSLNELALEPKDYGAVIVDVAQSKSAERLCNEDFNG 176
Qy 89 GLFSEDKVKYLNNE-----GLLSISEIEILVKMLFFFEKNKIDKEKVKYSRL 136
Db 177 SYFNSG-VVMYINREWLKQRLTEKFFDLSS-----DESIKKLKYPDQDILNLMFLHAKI 231
Qy 137 DPKSFGELYETLLHYDLRIADTTVHRIEDGVYL 170
Db 232 LPRKNCIYTIKSFBEKNSEYTRFINDTTFVI 265
XX
XX RESULT 6
XX ADW16234
XX ID ADW16234 standard; protein; 901 AA.
XX AC ADW16234;
XX DT 07-APR-2005 (first entry)
XX DE MmEI endonuclease protein.
XX KW Restriction endonuclease; CstMI endonuclease; methyltransferase;
XX KW genetic engineering; enzyme.
XX OS Unidentified.
XX
XX US2005009034-A1.
XX
XX 13-JAN-2005.
XX
XX 10-JUL-2003; 2003US-00616689.
XX
XX 10-JUL-2003; 2003US-00616689.
XX (MORG/) MORGAN R D.
XX (BHAT/) BHATIA T.
XX Morgan RD, Bhatia T;
XX WPI; 2005-074522/08.
XX
XX
XX New DNA and encoded Type II restriction endonuclease, CstMI, obtained
XX from Corynebacterium striatum, useful in genetic engineering or in
XX cleaving DNA at new positions within the DNA molecule.
XX Disclosure; SEQ ID NO 10; 28pp; English.
XX
XX The present invention relates to a novel DNA and encoded Type II
XX restriction endonuclease, CstMI or CstMI endonuclease/methyltransferase
XX obtained from Corynebacterium striatum. The invention is useful in
XX genetic engineering, particularly in cleaving DNA at new positions within
XX the DNA molecule. The present sequence is a MmEI endonuclease protein.
XX This sequence is used to align with CstMI protein
XX
XX Sequence 901 AA;
XX
XX Query Match 11.7%; Score 102; DB 9; Length 901;
XX Best Local Similarity 22.7%; Pred. No. 0.61;
XX Matches 34; Conservative 25; Mismatches 43; Indels 48; Gaps 6;
XX
Qy 16 IFILRIFFIAYIEDNDIFKKILOENKLYRSSIS-----FRYFFYDENTKKKLEYKKI 67
Db 182 LYLVRLFLCLPAEDTTIFEKSLFQEIETKLTLEDGSDLAHHINTLFYVLNTPQCKRLKN- 240
Qy 68 IITFNLLDKGSDAIFPVFNGSLFSE-----DKVKYLNNEGLLSISEIEILVKMLFF 120
Db 241 -----LDEHLAA--FPYINGKLFEEPLPPAQFDKAM---REALLDLCSLD----- 280
Qy 121 EKNKIDKEKVKYSRLDPKSFGELEYETLLE 150
Db 281 -----WSRISPAIFGSLFQSIMD 298
XX
XX RESULT 7
XX ADI61811
XX ID ADI61811 standard; protein; 919 AA.
XX AC ADI61811;
XX DT 22-APR-2004 (first entry)
XX DE MmEI restriction gene locus amino acid sequence SEQ ID NO:2.
XX KW MmEI; MmEI-like restriction enzyme; restriction enzyme; MmEI methylase;
XX KW endonuclease; methylase; enzyme; genetic engineering; genetic research;
XX KW cleavage.
XX OS Methylophilus methylotrophus.
XX WO2004007670-A2.
XX
XX 22-JAN-2004.
XX
XX 10-JUL-2003; 2003WO-US021570.
XX
XX 12-JUL-2002; 2002US-0395431P.
XX (NEWE ) NEW ENGLAND BIOLABS INC.
XX Morgan RD, Bhatia T, Davis T, Lovasco L;
XX WPI; 2004-122912/12.
XX N-PSDB; ADI61810.
XX
XX New MmEI or MmEI-like restriction endonucleases and their encoding DNAs,
XX useful in genetic engineering and research.
XX
XX Example 2; SEQ ID NO 2; 61pp; English.
XX
XX The present invention describes an isolated DNA (I) coding for the MmEI
XX or MmEI-like restriction enzyme. The DNA is obtainable from Methylophilus
XX methylotrophus. The DNA coding for MmEI-like restriction enzyme
XX hybridises to at least one conserved motif of the nucleotide sequence
XX coding for the MmEI restriction enzyme under predetermined conditions.

```

CC Also described: (1) a recombinant DNA vector comprising a vector into
CC which a DNA segment coding for the MmeI has been inserted, or comprising
CC the isolated DNA (1); (2) a host cell transformed by the vector of (1);
CC and (3) producing recombinant MmeI restriction endonuclease and MmeI
CC methylase, comprising culturing the host cell of (2) under conditions
CC suitable for expression of the endonuclease and methylase. The DNA and
CC the enzyme are useful in genetic engineering and in genetic research. The
CC enzymes allow precise cleavage of DNA at new positions within the DNA
CC molecule. The present sequence represents an amino acid sequence of the
CC MmeI restriction gene locus, which is used in the exemplification of the
CC present invention.
XX
SQ Sequence 919 AA;

Query Match 11.7%; Score 102; DB 8; Length 919;
Best Local Similarity 22.7%; Pred. No. 0.63;
Matches 34; Conservative 25; Mismatches 43; Indels 48; Gaps 6;
Qy 16 IFIIRIFIAIYENDIFKILQENKLYRSIS-----FRFFVDENTKKKLEYKKI 67
Dy 188 LYLVRLLFCLPAEDTTFEKSLEFQYETKTLEDGSLAHHINTLFYVLTPEQRLKN- 246
Qy 68 ITIFNLKDGSDATKFPVNGLFSE-----DKVKYLNNEGILLSISEIEILVKMLFF 120
Dy 247 -----LDEHLLA--FPYINGKLFEEPLPPAQQDKAM---REALDLCSLD----- 286
Qy 121 EERNIKDEKFKVYKSLRDPKSFGELEYTILE 150
Dy 287 -----WRSIPATFGSLFQSIMD 304

RESULT 8
ABU25689
ID ABU25689 standard; protein; 1032 AA.
XX
AC ABU25689;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #11216.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Clostridium difficile.
XX
PN WO200277193-A2.
XX
PD 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA29559.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 53613; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1032 AA;

Query Match 11.3%; Score 99; DB 6; Length 1032;
Best Local Similarity 27.5%; Pred. No. 1.5;
Matches 44; Conservative 28; Mismatches 50; Indels 38; Gaps 7;
Qy 34 KKILQENKLYRSISFRFFYDENTK---KKLEYKKIIT--IFNLDPKG--SDAIPFVF 86
Dy 88 KLIKWEGKHARLEIALDITEKEDISKISAEKLEIKELLVGCIKHILITESNFSDAVNLVLT 147
Qy 87 NGGLF-----SEDKVKYLN-----NEGIL-----LSISEIEILVKMLFFFEK- 123
Dy 148 NIGKFKADRAYVFEVSEDEKKTNTYEWNCNCGVDEQKALTNLDINELKNWFLLPQKG 207
Qy 124 -----NIKDEKFKVYKSLRDPKSFGELEYTILEYDLRI 155
Dy 208 FVIIDDIENIKESQPIEYELKPKIESLMTSSLKQDDKV 247

RESULT 9
AAE30333
ID AAE30333 standard; protein; 374 AA.
XX
AC AAE30333;
XX
DT 24-FEB-2003 (first entry)
XX
XX Human protein tyrosine phosphatase zeta SM1 variant protein.
XX
XX Brain; tumour protein target; Tbt; ischaemic stroke; cancer; epilepsy;
XX schizophrenia; depression; Alzheimer's disease; Parkinson's disease;
XX Huntington's chorea; traumatic head injury; dementia; stupor; headache;
XX coma; vertigo; weakness; myasthenia gravis; cerebrovascular disorder;
XX infection; multiple sclerosis; pregnancy; medical illness; vasotropic;
XX metabolic deficiency; cerebroprotective; antidepressant; antibacterial;
XX cytosstatic; nootropic; analgesic; fungicide; virucide; human; enzyme;
XX protein tyrosine phosphatase zeta SM1.
XX
XX Homo sapiens.
XX
XX WO200276510-A1.
XX
XX 03-OCT-2002.

XX 22-MAR-2002; 2002WO-US008992.
PF
XX
PR 23-MAR-2001; 2001US-00816703.
PR 17-OCT-2001; 2001US-00983000.
XX
XX (AGYT-) AGY THERAPEUTICS INC.
XX
XX Mueller S, Melcher T, Chin DJ;
XX
XX WPI; 2003-029903/02.
DR N-PSDB; AAD48126.
XX
XX
PT Developing active agents that modulate the activity of a brain tumor
PT protein target gene or gene product for treating e.g. stroke or cancer,
PT comprises contacting an agent with a brain tumor protein.
XX
XX Disclosure; Page 84-85; 135pp; English.
XX
XX The invention relates to a method for developing biologically active
XX agents that modulate activity of a brain tumor protein target (Tbt) gene
XX or gene product. The method is useful for developing biologically active
XX agents that modulate the activity of a brain tumor protein target gene
XX or gene product. Compounds that bind to the brain tumor proteins are
XX useful for treating e.g. ischaemic stroke, brain cancer, epilepsy,
XX schizophrenia, depression, Alzheimer's disease, Parkinson's disease,
XX Huntington's chorea, traumatic head injury, dementia, stupor, headache,
XX coma, vertigo, weakness, myasthenia gravis, cerebrovascular disorders,
XX infectious disorders (including fungal, bacterial, viral and parasitic
XX infections), multiple sclerosis, and other complications associated with
XX pregnancy, medical illness, alcohol and substance abuse, toxins and
XX metabolic deficiencies. The brain tumor proteins may also be used to
XX raise antibodies. The present sequence is human protein tyrosine
XX phosphatase zeta SM1 variant protein used to illustrate the method of the
XX invention
SQ Sequence 374 AA;

Query Match 11.1%; Score 97; DB 6; Length 374;
Best Local Similarity 21.8%; Pred. No. 0.6;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIENDIF---KKIQENKLYRSSIPRYFFYDENTKKLEYKII----- 68
Db 156 YCFDADRSSFEAVKGGKRALSLP-----EVGTEENLDFAKDGVESVRFKQA 210

QY 69 -----TIFNLKDGSDAIKPFVFNGLFS---EDKVKYLNNEGLLSISEIB-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPPTDVTVDWIVFKDVTVISESQLAVFCEVLT 268

QY 115 -----VKMLFPEEKNIKDEKFKYSR-----LDPKSFGE 143
Db 269 MQQSGYVLMYLDQNNFREQY-KFSRQVFSYTKESIHAVCSSPENVDADPENYTS 327

QY 144 L-----YETLLE-----VDLRIADTVVHRIEDG 167
Db 328 LLVTWERPRVYDTMIKFAVLYQQLDGEDQTKHEFLTDG 367

RESULT 10
ADL23826
ID ADL23826 standard; protein; 389 AA.
XX
AC ADL23826;
XX
DT 03-JUN-2004 (first entry)
XX
XX Pyrococcus furiosus intein Pfu IF2 amino acid sequence.
XX
XX cyclic peptide; interacting protein inhibition; fusion nucleic acid;
KW intein motif; intein; intein-catalysed cyclic peptide library;
KW cancer cell line; cardiovascular; neurobiology; bone biology;
KW biotechnology; skin biology; endocrinology; drug toxicity;

KW drug resistance; drug development.
XX
XX Pyrococcus furiosus.
PN WO2004019890-A2.
XX
XX 11-MAR-2004.
XX
XX 30-AUG-2003; 2003WO-US027371.
XX
XX 30-AUG-2002; 2002US-00232758.
PR 23-APR-2003; 2003US-00422536.
XX
XX (RIGE-) RIGEL PHARM INC.
XX
XX Lorens JB, Pray T, Kinsella TM, Bennett MK;
PI WPI; 2004-239115/22.
XX
XX
XX Generating intracellular cyclic peptide libraries, useful in identifying
XX interaction of interacting proteins by providing cells comprising first
XX and second interacting protein, and contacting cells with library of
XX fusion nucleic acids.
XX
XX Disclosure; Fig 3R; 156pp; English.
XX
XX The present invention describes a method (M1) for identifying a cell
XX comprising cyclic peptides which prevent interaction of interacting
XX proteins. M1 comprises providing cells with first and second interacting
XX proteins and contacting cells with a library of fusion nucleic acids
XX comprising (from 5' to 3'): a nucleic acid encoding a C-terminal intein
XX motif, a peptide, and an N-terminal intein motif, under conditions such
XX that cyclic peptide is formed. Also described is a method (M2) for
XX isolating the nucleic acid that encodes the cyclic peptide which when
XX introduced into cell caused the change in ubiquitination of the molecule,
XX comprising contacting a cell culture with an intein-catalysed cyclic
XX peptide library such that cells in the culture receive not more than one
XX member of the library, and monitoring the cells for a change in
XX ubiquitination of a molecule. M1 is useful for identifying a cell
XX comprising a cyclic peptide which is capable of preventing interaction of
XX interacting proteins. M1 is useful in creating cell lines from cancer
XX patients and in cardiovascular, neurobiology, drug toxicities and drug
XX biotechnology, skin biology, endocrinology, drug toxicities and drug
XX resistance applications. M1 is useful for identifying agents that
XX regulate intracellular and sarcolemmal calcium cycling in
XX cardiomyocytes, agents that diminish embolic phenomena in arteries and
XX arterioles, and for screening peptides that block transport mechanisms.
XX M1 is useful for screening enhanced contractile properties of
XX cardiomyocytes. M1 enables screening of a large number of cell types
XX under a wide variety of conditions. M1 enables improvement in the
XX performance of existing or developmental drugs. The present sequence is
XX used in the exemplification of the present invention.
SQ Sequence 389 AA;

Query Match 11.1%; Score 97; DB 8; Length 389;
Best Local Similarity 20.6%; Pred. No. 0.64;
Matches 45; Conservative 38; Mismatches 79; Indels 56; Gaps 7;

QY 5 HILASILEESIIFILRIFFIAYIEDNDIFKKIQENKLYRSSISF----- 49
Db 179 YIEVPLFEDPLLLIRFGIVSRIEKSTLKISGKNLELFRKHVGTDTSEKAKALDELISK 238

QY 50 -----RYFFVDE-----NTKKLE-----YKKIITIFNLLDKGSD--AIKFPYFN 87
Db 239 AKESERYPIIEELRRLGLLFGFTNRELRIEENPTYEIVIMEILERIERGSPNLAEKIAYLE 298

QY 88 GGLFSEDKVKYLNNEGLLSISEIEILVKMLFPEEKNIKDEKFKYSRLDPKSFGEVET 147
Db 299 GRIKEENYLRILEEGEGLIENGKLTGKLELVWNRNREFDSKQVDYVR-----NIVEN 351
QY 148 L-----LEYDLRIADTT--VHRIEDGVYLIRT 173

CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC neovascularisation syndromes, scarring and uterine fibroids. They may
CC also be useful in wound healing and in contraception. The present
CC sequence represents a polypeptide of the invention.
XX
SQ Sequence 1113 AA;
Query Match 11.1%; Score 97; DB 7; Length 1113;
Best Local Similarity 21.8%; Pred. No. 2.5;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;
Qy 26 YIEDNIF----KKILOENKLYRSSISFRFFYFDYNTKKKLEVKKII----- 68
Db 156 YCFDADRFSFEAEVKGKGLRALSILF-----EVGTEENLDKFAIIDGVESVSRFGKQA 210
Qy 69 -----TIFNLLDKGSDAIKPPVFNGLFS-----EDKVKYLNNEGLLSISEIE-----BIL- 114
Db 211 ALDPPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 268
Qy 115 -----VKMLPFEEKNIKDEKFKVYGR-----LDPKSFG 143
Db 269 MQQSGYVLMYLDLQNNFREQQY-KFSRQVFSYTGKEEIHAEVCSSEPNVQADPENYTS 327
Qy 144 L-----YETILE-----YDLRIADTTVHRIEDG 167
Db 328 LLVTWERPRVYDTMIEKFAVLYQQLDGEDQTKHEFLTGD 367
RESULT 13
ID ABUS6496
XX ABUS6496 standard; protein; 1337 AA.
AC ABUS6496;
XX
DT 02-APR-2003 (first entry)
XX Lung cancer-associated polypeptide #89.
XX
KW Lung cancer-associated polypeptide; cytostatic; emphysema;
KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
OS Unidentified.
XX
XX WO200286443-A2.
XX
XX 31-OCT-2002.
XX
XX 18-APR-2002; 2002WO-US012476.
XX
XX 18-APR-2001; 2001US-0284770P.
XX 10-MAY-2001; 2001US-0290492P.
XX 09-NOV-2001; 2001US-0339245P.
XX 13-NOV-2001; 2001US-0350666P.
XX 29-NOV-2001; 2001US-0334370P.
XX 12-APR-2002; 2002US-0372246P.
XX
XX (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Aziz N, Murray R;
XX
XX WPI; 2003-093161/08.
XX N-PSDB; ABX76220.
XX
XX Detecting a lung cancer-associated transcript in a cell from a patient
XX for treating lung cancer, by contacting a biological sample from the
XX patient with a polynucleotide that exhibits increased or decreased
XX expression in lung cancer.
XX
XX Claim 27; Page 262; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated
CC transcript in a cell from a patient, comprising contacting a biological
CC sample from the patient with a polynucleotide that selectively hybridises
CC to a sequence that is at least 80 % identical to a gene that exhibits
CC increased or decreased expression in lung cancer samples. Lung cancer-
CC associated polynucleotides and polypeptides are used for identifying a
CC compound that modulates a lung cancer-associated polypeptide, for
CC inhibiting proliferation of a lung cancer-associated cell to treat lung
CC cancer in a patient and for treating a mammal having lung cancer by
CC administering a modulatory compound identified. The methods are useful
CC for treating lung cancer, such as small cell lung cancer, non-small cell
CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC for diagnostic purposes and as targets for screening for therapeutic
CC compounds that modulate lung cancer, such as antibodies. Sequences
CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the
CC invention
XX
SQ Sequence 1337 AA;
Query Match 11.1%; Score 97; DB 6; Length 1337;
Best Local Similarity 21.8%; Pred. No. 3.2;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;
Qy 26 YIEDNIF----KKILOENKLYRSSISFRFFYFDYNTKKKLEVKKII----- 68
Db 38 YCFDADRFSFEAEVKGKGLRALSILF-----EVGTEENLDKFAIIDGVESVSRFGKQA 92
Qy 69 -----TIFNLLDKGSDAIKPPVFNGLFS-----EDKVKYLNNEGLLSISEIE-----BIL- 114
Db 93 ALDPPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 150
Qy 115 -----VKMLPFEEKNIKDEKFKVYGR-----LDPKSFG 143
Db 151 MQQSGYVLMYLDLQNNFREQQY-KFSRQVFSYTGKEEIHAEVCSSEPNVQADPENYTS 209
Qy 144 L-----YETILE-----YDLRIADTTVHRIEDG 167
Db 210 LLVTWERPRVYDTMIEKFAVLYQQLDGEDQTKHEFLTGD 249
RESULT 14
ABUS6690
ID ABUS6690 standard; protein; 1337 AA.
XX
XX AC ABUS6690;
XX
XX 02-APR-2003 (first entry)
XX
XX Lung cancer-associated polypeptide #283.
XX
XX Lung cancer-associated polypeptide; cytostatic; emphysema;
XX antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
XX small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
XX
XX Unidentified.
XX
XX WO200286443-A2.
XX
XX 31-OCT-2002.
XX
XX 18-APR-2002; 2002WO-US012476.
XX
XX 18-APR-2001; 2001US-0284770P.
XX 10-MAY-2001; 2001US-0290492P.
XX 09-NOV-2001; 2001US-0339245P.
XX 13-NOV-2001; 2001US-0350666P.
XX 29-NOV-2001; 2001US-0334370P.

KW	12-APR-2002; 2002US-0372246P.
XX	(BOSB-) EOS BIOTECHNOLOGY INC.
PI	Aziz N, Murray R;
XX	WPI; 2003-093161/08.
DR	N-PSDB; ABX76419.
XX	Detecting a lung cancer-associated transcript in a cell from a patient
PT	for treating lung cancer, by contacting a biological sample from the
PT	patient with a polynucleotide that exhibits increased or decreased
PT	expression in lung cancer.
XX	Claim 27; Page 408; 453pp; English.
XX	The invention relates to a method for detecting a lung cancer-associated
CC	transcript in a cell from a patient, comprising contacting a biological
CC	sample from the patient with a polynucleotide that selectively hybridises
CC	to a sequence that is at least 80 % identical to a gene that exhibits
CC	increased or decreased expression in lung cancer samples. Lung cancer-
CC	associated polynucleotides and polypeptides are used for identifying a
CC	compound that modulates a lung cancer-associated polypeptide, for
CC	inhibiting proliferation of a lung cancer-associated cell to treat lung
CC	cancer in a patient and for treating a mammal having lung cancer by
CC	administering a modulatory compound identified. The methods are useful
CC	for treating lung cancer, such as small cell lung cancer, non-small cell
CC	lung cancer or other benign or precancerous lesions, e.g. atelectasis,
CC	emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
CC	hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
CC	bronchiectasis. The genes, polynucleotides and polypeptides are useful
CC	for diagnostic purposes and as targets for screening for therapeutic
CC	compounds that modulate lung cancer, such as antibodies. Sequences
CC	ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
CC	cancer invention
XX	Sequence 1337 AA;
SQ	
Query Match	11.1%; Score 97; DB 6; Length 1337;
Best Local Similarity	21.8%; Pred. NO. 3.2;
Matches	48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;
QY	26 YIEDNDIF----KKILQENKLYRSSISFRYPFYDENTKKLEYKKII----- 68
DB	38 YCFDADRFSFBEAVKGKGLRAUSILF-----EVGTENLDPKAIDGVSVSRFGKQA 92
QY	69 -----TIFNLLDKGSDAIKFPVFNGLSPS---EDVKYLNNNEGLLSISEIE-----ETL- 114
DB	93 ALDPFFILLMLLNSTDT--KYIYINGSLTSPPCTDTVDWIVFDVTVSISESQLAVFCEVLIT 150
QY	115 -----VKMLPFEEKNIKDEKFKVYSR-----LDPKSGFE 143
DB	151 MQQSGYVLMMDYLQNFRQQY-KFSRQVFSSYTCKEIEHVAVCSEPEVNQADPENYTS 209
QY	144 L-----YETLE-----YDLRIADTTVTHRIEDG 167
DB	210 LLWTWRPRVVYDTMTKEPAVLYQLDGEDQTKEHFLTDG 249
RESULT 15	
ADN39060	
ID	ADN39060 standard; protein; 1337 AA.
XX	ADN39060;
XX	
DT	17-JUN-2004 (first entry)
DE	Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:378.
XX	
KW	Human; differential expression; cancer; angiogenic disorder;
KW	fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;
KW	inflammatory disease; autoimmune disease;
KW	retinal neovascularisation syndrome; scarring; uterine fibroid;

KW	detection; diagnosis; prognosis; drug screening; drug targeting;
KW	wound healing; contraception; cytostatic; cardiant; immunomodulatory;
XX	vulnerary; gene therapy; vaccine.
OS	Homo sapiens.
XX	WO2003042661-A2.
XX	22-MAY-2003.
XX	13-NOV-2002; 2002WO-US036810.
PR	13-NOV-2001; 2001US-0350666P.
PR	21-NOV-2001; 2001US-033464P.
PR	29-NOV-2001; 2001US-033493P.
PR	03-DEC-2001; 2001US-0335394P.
PR	14-DEC-2001; 2001US-0340376P.
PR	08-JAN-2002; 2002US-0347211P.
PR	10-JAN-2002; 2002US-0347349P.
PR	08-FEB-2002; 2002US-0355250P.
PR	13-FEB-2002; 2002US-0356714P.
PR	20-FEB-2002; 2002US-0359077P.
PR	03-MAR-2002; 2002US-0368809P.
PR	04-APR-2002; 2002US-0370110P.
PR	12-APR-2002; 2002US-0372248P.
PR	05-JUN-2002; 2002US-0386614P.
PR	16-JUL-2002; 2002US-0396839P.
PR	22-JUL-2002; 2002US-039775P.
PR	23-JUL-2002; 2002US-0397845P.
PR	03-SEP-2002; 2002US-0409450P.
XX	(BOSB-) EOS BIOTECHNOLOGY INC.
PA	
XX	Afar D, Aziz N, Ginsburg WM, Gish KC, Glynnne R, Hevezi PA;
PI	Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
XX	WPI; 2003-468649/44.
DR	N-PSDB; ADN39059.
XX	Determining the presence or absence of a pathological cell in a patient,
PT	useful for diagnosing, prognosing or treating cancer, comprises detecting
PT	a nucleic acid in a biological sample.
XX	Claim 12; SEQ ID NO 378; 1385pp; English.
CC	The invention relates to nucleic acids and proteins (ADN38683-ADN40064)
CC	whose expression is upregulated or downregulated in specific cancers or
CC	other diseases such as angiogenic or fibrotic disorders, and to methods
CC	of determining the presence or absence of a pathological cell in a
CC	patient by detecting a nucleic acid at least 80% identical to those of
CC	the invention or by detecting a polypeptide of the invention. The
CC	invention also relates to expression vectors and host cells comprising a
CC	nucleic acid of the invention; antibodies which specifically bind a
CC	polypeptide of the invention; use of such antibodies for drug targeting;
CC	and methods of screening for modulators of activity or expression of the
CC	polypeptides and nucleic acids. The nucleic acids, polypeptides,
CC	antibodies and methods are useful for diagnosing, prognosing and treating
CC	cancer and other conditions such as psoriasis, ischaemia, heart disease,
CC	atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC	neovascularisation syndromes, scarring and uterine fibroids. They may
CC	also be useful in wound healing and in contraception. The present
CC	sequence represents a polypeptide of the invention.
XX	Sequence 1337 AA;
SQ	
Query Match	11.1%; Score 97; DB 7; Length 1337;
Best Local Similarity	21.8%; Pred. NO. 3.2;
Matches	48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;
QY	26 YIEDNDIF----KKILQENKLYRSSISFRYPFYDENTKKLEYKKII----- 68
DB	38 YCFDADRFSFBEAVKGKGLRAUSILF-----EVGTENLDPKAIDGVSVSRFGKQA 92

QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGILLSISEIE-----EIL- 114
Db 93 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQSLAVFCEVLT 150
QY 115 -----VKMLFFBEKNIKDEKFKYISR-----LDPKSFG 143
Db 151 MQQSGYVMLMDYLQNNFREQQY-KFSRQVFSSYTGKEIHEAVCSSEPENNVQADPENYTS 209
QY 144 L-----YETILE-----YDLRIADTTVHRIIEDG 167
Db 210 LLVTWERPRVYDTMIEKFAVLYQQLDGEDQTKHEFLTGD 249

Search completed: January 24, 2006, 19:44:35
Job time : 40.8844 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 6.60107 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-14
Perfect score: 875
Sequence: 1 QTOHILASILESIIFILR.....IADTTVHRIEDGVYLIRTE 174

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872	99.7	1277	2 E70224	hypothetical prote
2	834	95.3	1278	2 B70236	hypothetical prote
3	232	26.5	1250	2 E81339	probable reestricti
4	182.5	20.9	706	2 T08313	conserved hypothet
5	104.5	11.9	337	2 F91191	probable LPS biosy
6	104.5	11.9	337	2 G86038	probable LPS biosy
7	104.5	11.9	867	2 B81419	probable type IIS
8	104.5	11.9	1084	2 C82931	hypothetical prote
9	101.5	11.6	609	2 B81385	probable ATP /GTP
10	99	11.3	273	2 C97189	glycosyltransferas
11	99	11.3	661	2 S67177	hypothetical prote
12	97	11.1	2314	1 A46151	protein-tyrosine-p
13	97	11.1	3119	2 T18414	protein g377 - mal
14	96.5	11.0	1181	2 B64516	hypothetical prote
15	95	10.9	558	2 S38614	hypothetical prote
16	95	10.9	744	2 G90124	hypothetical prote
17	94.5	10.8	408	2 E70127	flagellar motor sw
18	94	10.7	336	2 B90608	glycosyltransferas
19	94	10.7	1005	2 F90059	hypothetical prote
20	94	10.7	1044	2 H71049	translation initia
21	93.5	10.7	300	2 F69793	site-specific reco
22	93.5	10.7	619	2 B90141	hypothetical prote
23	93.5	10.7	858	1 A42239	adenylate cyclase
24	92.5	10.6	389	2 A70172	methyl-accepting c
25	92	10.5	806	2 AD0351	probable exported
26	91.5	10.5	223	2 G90126	hypothetical prote
27	91.5	10.5	454	2 I40861	iota toxin compone
28	91.5	10.5	610	2 A81365	ClaB protein Cj091
29	91	10.4	717	2 S78177	hypothetical prote

30	91	10.4	879	2 E69792	conserved hypothet
31	90.5	10.3	1306	2 T28313	ORF MSV152 probabl
32	89.5	10.2	136	1 ZNBPT4	dena protein - pha
33	89.5	10.2	395	2 F90012	hypothetical prote
34	89.5	10.2	564	2 S15962	hypothetical prote
35	89.5	10.2	2493	2 T40540	hypothetical prote
36	89	10.2	283	2 D97053	shape-determining
37	89	10.2	312	2 D84956	hypothetical prote
38	89	10.2	483	2 I40055	positive trans-act
39	89	10.2	763	2 A43747	X-Pro dipeptidyl-p
40	89	10.2	971	2 T70128	conserved hypothet
41	88.5	10.1	525	2 T28306	ORF MSV145 hypoth
42	88.5	10.1	591	2 D90592	hypothetical prote
43	88.5	10.1	808	2 E64492	hypothetical prote
44	88.5	10.1	1019	2 E90097	hypothetical prote
45	88.5	10.1	1817	2 H71611	probable secreted

ALIGNMENTS

RESULT 1

E70224
hypothetical protein BBE02 - Lyme disease spirochete plasmid E/lp25
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: E70224
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kervage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: E70224
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1277 <KLR>
A;Cross-references: UNIPROT:O50698; UNIPARC:UPI0000056833; GB:AE000785; NID:G2689951; P
A;Experimental source: strain B31
C;Genetics:
A;Genome: plasmid

Query Match	99.7%	Score 872;	DB 2;	Length 1277;
Best Local Similarity	99.4%	Pred. No. 2.6e-57;		
Matches 173;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	QTOHILASILESIIFILRIFIAIYIEDNDIFKKILOENKLYRSSISPRFFYDENTKK	60	
Db	268	EITQHTLASILESIIFILRIFIAIYIEDNDIFKKILOENKLYRSSISPRFFYDENTKK	327	
QY	61	KLEYKXIIITFNLDKSDAIKFPVNGGLFSEDKVKYLNNEGLLSISEIEELVKMLFF	120	
Db	328	KLEYKXIIITFNLDKSDAIKFPVNGGLFSEDKVKYLNNEGLLSISEIEELVKMLFF	387	
QY	121	EENKIDKFKVYSRLDPKSPGELYETLLEYDLRIADTTVHRIEDGVYLIRTE	174	
Db	388	EENKIDKFKVYSRLDPKSPGELYETLLEYDLRIADTTVHRIEDGVYLIRTE	441	

RESULT 2

B70236
hypothetical protein BBH09 - Lyme disease spirochete plasmid H/lp28-3
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C;Accession: B70236
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kervage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685

A:Accession: B70236
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1278 <KLE>
A:Cross-references: UNIPROT:O50667; UNIPARC:UPI0000056814; GB:AE000784; NID:g2690041; P
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 95.3%; Score 834; DB 2; Length 1278;
Best Local Similarity 92.5%; Pred. No. 1.8e-54;
Matches 161; Conservative 11; Mismatches 2; Indels 0; Gaps 0;

QY 1 QITQHILASLESIIFILRIFFIAYIEDNDIFPKILOENKLYRSSISPRYPYDENTKK 60
DB 274 EITRHLASLESIIFILRIFFIAYIEDNDIFPKILEENKLYRSSVSFRYPYDENTKK 333

QY 61 KLEYKKIITIFNLDDKGSDAIKFPVNGGLFSDKVKYLNNEGLLSISEIEILVKMLFF 120
DB 334 KLEYKKIITIFNLDDKGSDAIKFPVNGGLFQAQDKVKYLNNEGLLSISEIEILVKILFF 393

QY 121 EERNKDEKVKYSLRDPKSGELYETLLBYDRIADTTVHRIEDGVILIRTE 174
DB 394 EERNKDEKVKYSLRDPKSGELYETLLBYDRIADTTVHRIEDGVILIRTE 447

RESULT 3
E81339
probable restriction/modification enzyme Cj0690c [imported] - Campylobacter jejuni (str
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: E81339
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: E81339
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1250 <PAR>
A:Cross-references: UNIPROT:Q9PPL7; UNIPARC:UPI00000CICEE; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0690c

Query Match 26.5%; Score 232; DB 2; Length 1250;
Best Local Similarity 33.7%; Pred. No. 1.4e-09;
Matches 55; Conservative 34; Mismatches 68; Indels 6; Gaps 3;

QY 3 TQHILASLESIIFILRIFFIAYIEDNDIFPKILOENKLYRSSISPRYPF--YDENTKK 60
DB 260 TKADLKLIYENSIFYIFRLFIAYFEDK--FEILEKHKVFYKSKISLTLENLQDESS 317

QY 61 KLEYKKIITIFNLDDKGSDAIKFPVNGGLFSDKVKYLNNEGLLSISEIEILVKMLFF 120
DB 318 SGGEGELENIYNYKKGKGFMDVPVNGGLFDESKTALLSTPKIFNDKDLKFIQLNLF 377

QY 121 EERNKDEKVKYSLRDPKSGELYETLLBYDRIADTTVHRI 163
DB 378 KDKNLSFKR--DYKTLVSEHLGTIVEGLLSYFPEIANEDIVYV 418

RESULT 4
T08313
conserved hypothetical protein H1130 - Halobacterium sp. (strain NRC-1) plasmid pNRC100
C:Species: Halobacterium sp.
A:Variety: strain NRC-1
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08313
R:Ng, W.V.; Clufo, S.A.; Smith, T.M.; Bungarner, R.E.; Baekin, D.; Faust, J.; Hall, B.;
Genome Res. 8, 1131-1141, 1998
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or m

A:Reference number: Z16408; MUID:99063795; PMID:9847077
A:Accession: T08313
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-706 <DAS>
A:Cross-references: UNIPROT:O52000; UNIPARC:UPI00000631E7; EMBL:AF016485; NID:g2822278;
A:Experimental source: strain NRC-1
C:Genetics:
A:Gene: HALOSP:H1130
A:Genome: plasmid pNRC100

Query Match 20.9%; Score 182.5; DB 2; Length 706;
Best Local Similarity 27.6%; Pred. No. 3.6e-06;
Matches 55; Conservative 39; Mismatches 58; Indels 47; Gaps 10;

QY 10 ILEESIIFILRIFFIAYIEDNDIFPKILO-ENKLYRSSISF-----RYF 52
DB 274 IHDSLIYLRILFVLYAEAG--RELDTSNEIYEQSYSLNSLKEIAEELDSGDPKVR 331

QY 53 FYDENTKKLYKKIITIFNLDDKGS-----DAIKFPVNGGLFSED-----KVY 98
DB 332 DQDNLOSLD-----ELFTLINKSKSGRIGIPEEDLYIPAYNGGLFRTDPDEDSAEAKF 386

QY 99 INNE--GLLSISEIEILVKMLFFEEKNIKDEK-FVKYSLRDPKSGELYETLLYDLRI 155
DB 387 LANHDVGDAYLAKWELLTR----SKNDGGGKIFVDYSSLDVRHLGSIYEGLEQLNV 442

QY 156 ADTTHRIEDGVILIRTE 174
DB 443 ADEPL--ALDDGEYVSADE 459

RESULT 5
F91191
probable LPS biosynthesis enzyme waaJ [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F91191
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91191
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <HAY>
A:Cross-references: UNIPROT:Q8XDC2; UNIPARC:UPI00000D0396; GB:BA000007; PIDN:BA837925.1
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs4502
C:Superfamily: lipopolysaccharide glucosyltransferase I

Query Match 11.9%; Score 104.5; DB 2; Length 337;
Best Local Similarity 21.0%; Pred. No. 1.1;
Matches 45; Conservative 46; Mismatches 74; Indels 49; Gaps 8;

QY 1 QITQHILASLESIIFIL-----RIFIAIYENDNIFKKILOEN---KLYRSSISF 49
DB 57 KFTFHIADYLDKEYIELLSQLATKYQTVIKLYHIDSEPLKALPQSNIPVSIYRLLSF 116

QY 50 RYF-----FYDENTKKK-----LEYKK-----IITIFNLDDKGSDAIKFPVNG 88
DB 117 DYFSARLDLSLYLDADIVCKGSLNELIALEFKDEYAGNVIDVDMQSKSAERLCNEDFNG 176

QY 89 GLFSEDKVKYLNNE-----GLLSISEIEILVKMLFFEEKNIKDEKPKVKYSRL 136
DB 177 SYFNSG-VWYNLRLEWMLKQRLTEKFFDLS-----DESIKKLYKYPQDILNMLFHHAKI 231

QY 137 DPKSFGELYETLLBYDRIADTTVHRIEDGVIL 170
DB 232 LPRKYNCIYTIKSEFEKSEKSEYTRFINDDTVFI 265

Best Local Similarity 28.0%; Pred. No. 3.4;
Matches 47; Conservative 21; Mismatches 59; Indels 41; Gaps 8;

QY 1 QITQHILASTLEESIIIFILRIPFIATIEDNDIFKKILOENKLYR-----SSISFRFYFFVD 55
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
347 RISSEIPASVKEKDVP-----YKESGFLUKOLYTRYDYKAPYISDDAFLAMFTN 398
QY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
56 ENTKKKLEYKKIITIFNLILDKGSDAIKFPVNGGLFSEDKVYLNNEGLLSISEIBEILV 115
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
399 SDVMSK-EPKK---IKNELYSKFEEIKMKL-----KOPINILEREILLFKAFSNSIQK 447
QY 116 KMLFFBEKNI-----KDEPKVKYSRLDPKSPGE-LYETLLLEVDL 153
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
448 DHIFQSKNPSSELRAFNCNASDEYFL-----KDFKELLFKSILELDL 488

RESULT 10
C97189
glycosyltransferase [imported] - Clostridium acetobutylicum
C/Species: Clostridium acetobutylicum
C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C/Accession: C97189
R/Molling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A/Reference number: A96900; MUID:21359325; PMID:21359325
A/Accession: C97189
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-273 <KUR>
A/Cross-references: UNIPROT:Q97GM0; UNIPARC:UIP00000CA4C2; GB:AEO01437; PIDN:AAK80302.1;
A/Experimental source: Clostridium acetobutylicum ATCC824
C/Genetics:
A/Gene: CAC2346

Query Match 11.3%; Score 99; DB 2; Length 273;
Best Local Similarity 25.9%; Pred. No. 2.2;
Matches 44; Conservative 30; Mismatches 50; Indels 46; Gaps 10;

QY 23 FIAYIEDNDIFKKILOENKLYRSISPRYPFYDENTKKLEYKKIITIF---NLLDKGSD 79
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
90 YINYLNDDDMFHK---SKIMR---MIRYF-----QNYKKITLVTSHRSVIDKNDN 133
QY 80 AIKFPVNGGLFSDKVKYLNNEGLLSISIEIELVKMLFEEKNIKDEKFKYSRLDPK 139
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
134 IL--PDIN----ATKRIRYLNR-----EVS--GRIAKLMFNLGNFIGEPTTVLFKKSVDK 181
QY 140 SFGEL-----YETILE-----YDLRIADTIVHRIIEDGVYLIRTE 174
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
182 NFGSLNGKYQLYSLIDAATWTFELLSKGYCVVIADTLTSYFRHEGQNTWRPE 231

RESULT 11
S67177
hypothetical protein YOR275c - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein O5450
C/Species: Saccharomyces cerevisiae
C/Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C/Accession: S67177; S72046
R/Cheret, G.; Sor, F.
submitted to the Protein Sequence Database, July 1996
A/Reference number: S67169
A/Accession: S67177
A/Molecule type: DNA
A/Residues: 1-661 <CHE>
A/Cross-references: UNIPROT:Q12033; UNIPARC:UIP000006C82B; EMBL:Z75183; NID:g1420615; P
A/Experimental sources: strain S288C
R/Cheret, G.; Bernardi, A.; Sor, F.
Yeast 12, 1059-1064, 1996
A/Title: DNA sequence analysis of the VP1-SNF2 region on chromosome XV of Saccharomyc
A/Reference number: S72039; MUID:97051594; PMID:8896271
A/Accession: S72046

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-661 <HW>

A;Cross-references: UNIPARC:UPI000006C82B; EMBL:X89633; NID:g1279694; PIDN:CAA61781.1; A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995

C;Genetics:

A;Cross-references: SGD:S0005801

A;Map position: 15R

A;Note: YOR275c

Query Match 11.3%; Score 99; DB 2; Length 661;
Best Local Similarity 30.8%; Pred.No. 5.6; Gaps 11;
Matches 45; Conservative 22; Mismatches 37; Indels 42; Gaps 11;

Qy 28 EDND----IPKTIQNLKLYRSSISFRYFFYDENTKKCLEYK---KIITIFNLLDKGSDA 80
Db |||||TNNKTKMIQEQ-----RNDYIDTRKSSSEYRILPKIITSY-----KKNGT 537
Qy 81 IKF-PVFNGL--FSDKVKYLNNEGLLSISETEERILVLMFPFEKN-----IKDEKFFV 131
Db :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 538 VDEPIFIGHLKYFDED-LRYVNSTKEENIKLIEEYVNL-----SKGNPGRSGIEPKQWV 591
Qy 132 KYSELDPKSGELYETLLVLDLIA 157
Db ||||:|||||----ELYIEDLRYSFKLLD 611
Qy 592 ---RIDPR---ELYIEDLRYSFKLLD 611

RESULT 12

A46151

A;Title: protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type zeta precursor - human

N;Alternate names: protein-tyrosine-phosphatase beta; protein-tyrosine-phosphatase xi

C;Species: Homo sapiens (man)

C;Date: 08-Dec-1993 #sequence revision 08-Feb-1996 #text_change 31-Dec-2004

C;Accession: A46151; A46700; B36065; S12054

R;Krueger, N.X.; Saito, H.
Proc. Natl. Acad. Sci. U.S.A. 89, 7417-7421, 1992

A;Title: A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is expressed in

A;Reference number: A46151; MUID:92366472; PMID:1323835

A;Accession: A46151

A;Molecule type: mRNA

A;Residues: 1-2314 <RU>

A;Cross-references: UNIPROT:P23471; UNIPARC:UPI00001329AB; GB:M93426; NID:g190743; PIDN

A;Experimental source: fetal brain

A;Note: sequence extracted from NCBI backbone (NCBIN:110851, NCBI:P:110852)

A;Note: sequence inconsistent with the nucleotide translation

R;Levy, J.B.; Canoll, P.D.; Silvennoinen, O.; Barnea, G.; Morse, B.; Honegger, A.M.; Hu
nger, J.
J. Biol. Chem. 269, 10573-10581, 1993

A;Title: The cloning of a receptor-type protein tyrosine phosphatase expressed in the c

A;Reference number: A46700; MUID:93252948; PMID:8387522

A;Accession: A46700

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-1721,1729-2314 <LEV>

A;Cross-references: UNIPARC:UPI0000173863

A;Experimental source: brainstem

A;Note: sequence extracted from NCBI backbone (NCBIP:131344)

R;Kaplan, R.; Morse, B.; Huebner, K.; Croce, C.; Croce, R.; Ravera, M.; Ricca, G.; Jaye,
Proc. Natl. Acad. Sci. U.S.A. 87, 7000-7004, 1990

A;Title: Cloning of three human tyrosine phosphatases reveals a multigene family of rec

A;Reference number: A36065; MUID:90384936; PMID:2169617

A;Accession: B36065

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1749-1990;2047-2280 <XAP>

A;Cross-references: UNIPARC:UPI000014423B; UNIPARC:UPI000014423B; GB:M34668; NID:g19073

R;Krueger, N.X.; Streuli, M.; Saito, H.
EMBO J. 9, 3241-3252, 1990

A;Title: Structural diversity and evolution of human receptor-like protein tyrosine phs

A;Reference number: S12049; MUID:91006018; PMID:2170109

A;Accession: S12054

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1479-2091 <KR2>
A;Cross-references: UNIPARC:UPI000016AF42; GB:X54135; NID:935795; PIDN:CAA38070.1; PID:9
C;Genetics:
A;Gene: GDB:PTPR21; PTP2; PTP2; PTP2; PTP18; RPTPB
A;Cross-references: GDB:127353; OMIM:176891
A;Map position: 7q31.3-7q31.3
C;Function:
A;Description: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate
A;Note: may be involved in the regulation of specific developmental processes in the cen
C;Superfamily: protein-tyrosine-phosphatase, receptor type beta; carbonic anhydrase hom
C;Keywords: alternative splicing; brain; glycoprotein; phosphoprotein; phosphoric mono
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1635/Domain: protein-tyrosine-phosphatase, receptor type zeta #status predicted <M
F;25-1635/Domain: extracellular #status predicted <EXT>
F;25-754,1614-2314/Product: protein-tyrosine phosphatase receptor type zeta, short form
F;38-300/Domain: carbonic anhydrase homology <CAH>
F;1636-1661/Domain: transmembrane #status predicted <TMN>
F;1662-2314/Domain: intracellular #status predicted <INT>
F;1749-1980/Domain: protein-tyrosine-phosphatase homology <PTP1>
F;1047-2270/Domain: protein-tyrosine-phosphatase homology <PTP2>
F;105,134,223,232,324,381,480,497,501,552,602,629,677,782,1017,1050,1082,1122,1456,1561,
F;1932/Active site: Cys (phosphocysteine intermediate) #status predicted
F;1938/Binding site: substrate phosphate (Arg) #status predicted

Query Match 11.1%; Score 97; DB 1; Length 2314;
Best Local Similarity 21.8%; Pred. No. 31;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDF-----KKIQENKLYRSSISFRFFYDENTKKLYKKII----- 68
Db 156 YCFDADFSSFEAVKGGKRLALSILF-----EVGTEENLDFKAIIDGVESVSRFGKQA 210
QY 69 -----TIFNLLDGSDAIKFPVNGGLFS---EDKVKYLNNEGILLSIEIE-----EIL- 114
Db 211 ALDPFILLNLPSTD---KYIYNGSLTSPCTDVTDMWIFKDTVSISQLAVFCEVLT 268
QY 115 -----VKMLPFEEKIKDEKFKVYSR-----LDPKSFGE 143
Db 269 MQQSGVYMLDYQNNFREQQY-KFSRQVSSVYTGKEEIHAVCSSEPNVQADPENYTS 327
QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 328 LLVTWERPVVYDTMEKFAVLVQQLDGEDQTKHEPLTDG 367

RESULT 13
T18414
protein g377 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18414
R;Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Cappai, R.
Mol. Biochem. Parasitol. 74, 143-156, 1995
A;Title: The Leishmania promastigote surface antigen 2 complex is differentially expres
A;Reference number: Z18933; MUID:96360472; PMID:8719156
A;Accession: T18414
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3119 <HAN>
A;Cross-references: UNIPROT:Q25857; UNIPARC:UPI0000082072; EMBL:L04161; NID:9309687; PID

Query Match 11.1%; Score 97; DB 2; Length 3119;
Best Local Similarity 25.7%; Pred. No. 43;
Matches 38; Conservative 23; Mismatches 41; Indels 46; Gaps 7;

QY 18 ILRIFVIAYIEDNDIFKKIQENKLYRSSISFRFFYDENTKKLYKKIIITFNLLDKG 77
Db 2599 IIEPTTKFTIENKPVFKNM---PLLXNRRIKY-----DSINEKLOYKEELDLDHMDNE 2649
QY 78 SDAIKFPVNGGLFSDDKVKYLNNEGILLSIEIEILVKMLFF-----EEKNKID--- 127
Db 2650 S-----FHQ-----LLDLINKESITLQSLFYSRYTEHANIKDFPK 2687

QY 128 -----EKFKVYSRLDPKPSFGELYETILLEY 151
Db 2688 LPMFKKXYTVQ--DEKFFINLYEYLYNY 2714

RESULT 14
B64516
hypothetical protein MJEC502 - Methanococcus jannaschii plasmid PURB801
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: B64516
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschi
A;Reference number: A64300; MUID:96337999; PMID:8688087
A;Accession: B64516
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1181 <BUL>
A;Cross-references: UNIPROT:Q60301; UNIPARC:UPI000013C258; GB:L771119; NID:91500688; TIG
C;Genetics:
A;Map position: ECSREV4814-1269
A;Genome: plasmid
A;Start codon: GTG
A;Note: this stable 16-kilobase pair plasmid is also designated ECS (small extrachromos

Query Match 11.0%; Score 96.5; DB 2; Length 1181;
Best Local Similarity 22.4%; Pred. No. 16;
Matches 36; Conservative 32; Mismatches 46; Indels 47; Gaps 8;

QY 14 SIIFILRIFFIAYIED-----NDIFKKIQENKLYRSSISFRFFY-----FYD----- 55
Db 290 AVLLMNLRIFFILEDKGIQVPRDLRLRYEDYK--KSNVLINVDAYLKPFLFEVLNTP 347
QY 56 -----ENTKKLYKKIIITFNLLDGSDAIKFPVNGGLFSDDKVKYLNNEGILLSIEIE 111
Db 348 DERKENIRTNPYKDI-----PYLNGGLFRSNVNP---NELSFTIKD-N 387
QY 112 EILVKMLFEEK---NIKDEKFKVYSRLDPKPSFGELYETIL 149
Db 388 EIIGVINFLERYKFTLSTSGSEVEVLNPDILGYVYEKLI 428

RESULT 15
S38614
hypothetical protein 558 (tufA 3' region) - euglenid (Astasia longa) plastid
C;Species: plastid Astasia longa
C;Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
C;Accession: S38614
R;Gockel, G.; Baier, S.; Hachtel, W.
submitted to the EMBL Data Library, November 1993
A;Reference number: S38590
A;Accession: S38614
A;Molecule type: DNA
A;Residues: 1-558 <GOC>
A;Cross-references: UNIPROT:P34782; UNIPARC:UPI000013A77B; EMBL:X75652; NID:9414919; PI
C;Keywords: chloroplast; plastid

Query Match 10.9%; Score 95; DB 2; Length 558;
Best Local Similarity 23.2%; Pred. No. 9.3;
Matches 39; Conservative 34; Mismatches 59; Indels 36; Gaps 7;

QY 22 FFIAYIEDNDIFKKIQENKLYRSSISFRFFYDENTKKLYKKIIITFNLLDK--GSD 79
Db 72 YFIINLEDFKFIETFTDSVLIVSGVTRFFHSESEQKDS-----LIFNFIPIEINSD 136
QY 80 AIKFPVNGGLFSDDKVKYLNNEGILLSIEIEILVKMLFF-----LNNEGILLSIEIEILVKMLFF 120
Db 127 GL---MGTKVLLESKILYENENKEKDLKLVHFIPTDIINNVLIGVKVLE---KKKIFY 181

Qy 121 EEKNIKDKFVKYSRLDPKSF--GELYETLLEYDLRIADTIVHRIED 166
Db 182 ENENIKOLEFIEGSELEGGQEIKSEIEEKVVDKY---DDSEGRVLLD 226

Search completed: January 24, 2006, 19:59:06
Job time : 8.60107 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 37.3889 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-14
Perfect score: 875
Sequence: 1 QITQHILASILESIIFILR.....IADTTVHRIEDGVYLIRTE 174

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872	99.7	1277	2	O50698 BORBU
2	834	95.3	1278	2	O50667 BORBU
3	821	93.8	1194	2	O5XYT2_BORGA
4	470	53.7	149	2	O5XYI6_BORGA
5	290	33.1	806	2	O5XYG5_BORGA
6	232	26.5	1250	2	O5HYV1_CAMJUE
7	232	26.5	1250	2	O5PPL7_CAMJUE
8	182.5	20.9	662	2	O9HI06_HALSA
9	182.5	20.9	706	2	O52000_HALSA
10	109.5	12.5	1193	2	O5ON37_ENTHI
11	109.5	12.5	2181	2	O51017_ENTHI
12	108.5	12.4	2337	2	O4XXL0_PLACH
13	107.5	12.3	868	2	O54H63_DICDI
14	107.5	12.3	939	2	O5FM50_LACAC
15	106.5	12.2	1524	2	O8FSZ4_COREF
16	106	12.1	1572	2	O4XYM9_PLACH
17	105.5	12.1	309	2	O834E5_ENTFA
18	105.5	12.1	837	2	O6F1B2_MESFL
19	105	12.0	634	2	O7R9S1_PLAYO
20	104.5	11.9	337	2	O83PN9_SHIFL
21	104.5	11.9	337	2	O8XDC2_ECOS7
22	104.5	11.9	867	2	O9PJ80_CAMJUE
23	104.5	11.9	878	2	O4HEH3_CAMCO
24	104.5	11.9	1084	2	O5PRL8_UREPA
25	104.5	11.9	1257	2	O5HXC7_CAMJUE
26	103	11.8	687	2	O4XNV9_PLACH
27	102.5	11.7	788	2	O7RIG0_PLAYO
28	102	11.7	1843	2	O8IC40_PLAF7
29	101.5	11.6	609	2	O9PI93_CAMJUE
30	101.5	11.6	2259	1	YCF2_PHYPA
31	100.5	11.5	723	2	O8D3I4_WIGBR

32	100.5	11.5	970	2	O8I3G3_PLAF7	Q8i9g3 plasmodium
33	100.5	11.5	1610	2	O5FQUB_GLUOX	Q5fqu8 gluconobact
34	100	11.4	856	2	O4Z0W0_PLABE	O4z0w0 plasmodium
35	100	11.4	1056	2	O7VII1_HELHP	O7vii1 helicobacte
36	99.5	11.4	258	2	O8I2E9_PLAF7	Q8i2e9 plasmodium
37	99.5	11.4	371	2	O4FNN9_9RICK	O4fnn9 candidatus
38	99.5	11.4	486	2	O4ZIT6_PLABE	O4zit6 plasmodium
39	99.5	11.4	3377	2	O8IEH5_PLAF7	Q8ieh5 plasmodium
40	99	11.3	273	2	O97GM0_CLOAB	Q97gm0 clostridium
41	99	11.3	290	2	O7NAL6_MYCGA	O7nal6 mycoplasma
42	99	11.3	661	1	PALA_YEAST	Q12033 saccharomyc
43	99	11.3	930	2	O6MAP0_PARUM	O6map0 parachlamyd
44	98	11.2	714	2	O4YHMS_PLABE	Q4yhm5 plasmodium
45	97.5	11.1	291	2	O4MUDI_BACCE	Q4mudi bacillus ce

ALIGNMENTS

RESULT 1

ID	O50698 BORBU	PRELIMINARY;	PRT; 1277 AA.
AC	O50698;		
DT	01-JUN-1998	(TREMREL. 06, Created)	
DT	01-JUN-1998	(TREMREL. 06, Last sequence update)	
DE	01-MAR-2004	(TREMREL. 26, Last annotation update)	
DE	Hypothetical protein BBE02.		
GN	OrderedLocusNames=BBE02;		
OS	Borrelia burgdorferi (Lyme disease spirochete).		
OG	Plasmid lp25.		
OC	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;		
OC	Borrelia burgdorferi group.		
OX	NCBI_TaxID=139;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=ATCC 35210 / B31;		
RX	MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;		
RA	Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,		
RA	Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,		
RA	Gwin M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,		
RA	Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,		
RA	Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,		
RA	Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,		
RA	Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,		
RA	Roberts K.M., Hatch B., Smith H.O., Venter J.C.;		
RT	"Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."		
RL	Nature 390:580-586(1997).		
DR	EMBL; AE000785; AAC66031.1; -; Genomic_DNA.		
DR	PIR; E70224; E70224.		
DR	HSSP; Q57997; 1MJH.		
DR	TIGR; BBE02; -		
DR	GO; GO:0003677; F.DNA binding; IEA.		
DR	GO; GO:0008170; F.N-methyltransferase activity; IEA.		
DR	GO; GO:0006306; P.DNA methylation; IEA.		
DR	InterPro; IPR007409; DUF450.		
DR	InterPro; IPR002236; N12N6_mtfase.		
DR	InterPro; IPR002052; N6_Mtase.		
DR	Pfam; PF04313; HSDR N; 1.		
DR	PRINTS; PR00507; N12N6MTFAS.		
DR	PROSITE; PS00092; N6_MTASE; UNKNOWN 1.		
KW	Complete proteome; Hypothetical protein; Plasmid.		
SQ	SEQUENCE 1277 AA; 150972 MW; 0841DF4EDB5859F4 CRC64;		

Query Match 99.7%; Score 872; DB 2; Length 1277;
Best Local Similarity 99.4%; Pred. No. 1.2e-55;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	QITQHILASILESIIFILRIFIAIYIENDIPFKILQENKLYRSISFYFFYDENTKK	60
DB	268	EITQHILASILESIIFILRIFIAIYIENDIPFKILQENKLYRSISFYFFYDENTKK	327
QY	61	KLEYKKIITFIFNLLDKSGDAIKFPVPVNGGLFSDKVKYLNNEGLLSISEIEILVKMLFF	120

```

Db      328 KLEYKKIITIFNLLDKGSDAIFKPFVNGGLFSEDKVKYLNNEGLLSISEIEELVVKMLFF 387
QY      121 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 174
Db      388 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 441

RESULT 2
OS0667 BORBU
ID      OS0667 BORBU PRELIMINARY; PRT; 1278 AA.
AC      OS0667;
RC      01-JUN-1998 (T-EMBLrel. 06, Created)
DT      01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT      01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE      Hypothetical protein BBH09.
GN      OrderedLocusNames=BBH09;
OC      Borrelia burgdorferi (Lyme disease spirochete).
OG      Plasmid lp28-3.
OC      Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC      Borrelia burgdorferi group.
OX      NCBI_TaxID=139;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=ATCC 35210 / B31;
RX      MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA      Fraser C.M., Caefjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA      Lathigra R., White O., Ketchum K.A., Dodson R.J., Hackey E.K.,
RA      Winn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA      Richardson D.L., Petersen J.D., Kerlavage A.R., Quackenbush J.,
RA      Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA      Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA      Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA      Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT      "Genomic sequence of a Lyme disease spirochete, Borrelia
RT      burgdorferi.";
RL      Nature 390:580-586 (1997).
DR      EMBL; AE000784; AAC66000.1; -; Genomic_DNA.
DR      PIR; B70236; B70236.
DR      TIGR; BBH09; -.
DR      GO; GO:0003677; F:DNA binding; IEA.
DR      GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR      GO; GO:0006306; P:DNA methylation; IEA.
DR      InterPro; IPR007409; DUF450.
DR      InterPro; IPR002296; N12N6_mtfase.
DR      Pfam; PF04313; HSDR N; 1.
DR      PRINTS; PR00507; N12N6MTFRASE.
DR      PROSITE; PS00092; N6_MTASE; UNKNOWN 1.
KW      Complete proteome; Hypothetical protein; Plasmid.
SQ      SEQUENCE 1278 AA; 150871 MW; 8A0B22DF166565C0 CRC64;

Query Match 95.3%; Score 834; DB 2; Length 1278;
Best Local Similarity 92.5%; Pred. No. 7.4e-53;
Matches 161; Conservative 11; Mismatches 2; Indels 0; Gaps 0;

QY      1 QITQHILASILEESIIIFILRIFFIAYIEDNDIFKKILOENKLYRSSISFRYFFYDENTKK 60
Db      274 EITRHILASILEESIIIFILRIFFIAYIEDNDIFKKILEENKLYRSSVSFRYFFYDENTKK 333
QY      61 KLEYKKIITIFNLLDKGSDAIFKPFVNGGLFSEDKVKYLNNEGLLSISEIEELVVKMLFF 120
Db      334 KLGYYKKIITIFNLLDKGSDAIFKPFVNGGLFQAQDKVLYNNESLLSISEIEELVVKILFF 393
QY      121 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 174
Db      394 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 447

RESULT 3
Q5XYT2 BORGA
ID      Q5XYT2 BORGA PRELIMINARY; PRT; 1194 AA.
AC      Q5XYT2;

```

```

DT      25-OCT-2004 (T-EMBLrel. 28, Created)
DT      25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT      25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE      Hypothetical protein.
GN      ORFNames=BGP199;
OC      Borrelia garinii Pbi.
OC      Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC      Borrelia burgdorferi group.
OX      NCBI_TaxID=290434;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=Pbi;
RA      Glockner G., Lehmann R., Romualdi A., Pradella S.,
RA      Schulte-Spechtel U., Schilhabel M., Wilske B., Suhnel J., Platzner M.;
RT      "Comparative analysis of the Borrelia garinii genome.";
RL      Nucleic Acids Res. 32:6038-6046(2004).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=Pbi;
RA      Glockner G., Schilhabel M., Lehmann R., Platzner M.;
RL      Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY722925; AAU86050.1; -; Genomic_DNA.
DR      GO; GO:0003677; F:DNA binding; IEA.
DR      GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR      GO; GO:0006306; P:DNA methylation; IEA.
DR      InterPro; IPR007409; DUF450.
DR      InterPro; IPR002296; N12N6_mtfase.
DR      InterPro; IPR002052; N6_Mtase.
DR      Pfam; PF04313; HSDR N; 1.
DR      PRINTS; PR00507; N12N6MTFRASE.
DR      PROSITE; PS00092; N6_MTASE; UNKNOWN 1.
KW      Hypothetical protein.
SQ      SEQUENCE 1194 AA; 141642 MW; 51DC343D205F2E41 CRC64;

Query Match 93.8%; Score 821; DB 2; Length 1194;
Best Local Similarity 92.0%; Pred. No. 6.3e-52;
Matches 160; Conservative 10; Mismatches 4; Indels 0; Gaps 0;

QY      1 QITQHILASILEESIIIFILRIFFIAYIEDNDIFKKILOENKLYRSSISFRYFFYDENTKK 60
Db      268 EITQILASILEESIIIFILRIFFIAYIEDNDIFKKILEENKLYRSSISFRYFFYDENTKK 327
QY      61 KLEYKKIITIFNLLDKGSDAIFKPFVNGGLFSEDKVKYLNNEGLLSISEIEELVVKMLFF 120
Db      328 KLGYYKKIITIFNLLDKGSDAIFKPFVNGGLFAEDKVKYLNNEGLLSISEIEELVVKILFF 387
QY      121 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 174
Db      388 EERNKIDKFKVYSRLDPKPSFGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 441

RESULT 4
Q5XY16 BORGA
ID      Q5XY16 BORGA PRELIMINARY; PRT; 149 AA.
AC      Q5XY16;
RC      25-OCT-2004 (T-EMBLrel. 28, Created)
DT      25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT      25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE      Hypothetical protein.
GN      ORFNames=BGP295;
OC      Borrelia garinii Pbi.
OC      Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC      Borrelia burgdorferi group.
OX      NCBI_TaxID=290434;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=Pbi;
RA      Glockner G., Lehmann R., Romualdi A., Pradella S.,
RA      Schulte-Spechtel U., Schilhabel M., Wilske B., Suhnel J., Platzner M.;
RT      "Comparative analysis of the Borrelia garinii genome.";
RL      Nucleic Acids Res. 32:6038-6046(2004).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.

```


Q5HV91;
10-MAY-2005 (TrEMBLrel. 30, Created)
10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
Hypothetical protein.
OrderedLocusNames=CJ06789;
OS Campylobacter jejuni (strain RM1221).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
NCBI_TaxID=195099;
[1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP PubMed:15660156; DOI=10.1371/Journal.pbio.0030015;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Rasko D.A.,
RA Ravel J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Shvartsbeyn A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural differences and novel potential virulence mechanisms
RT from the genomes of multiple Campylobacter species.";
RL PLOS Biol. 3:72-85(2005).
RW EMBL: CP000025; AAW34576.1; -- Genomic_DNA.
DR TIGR; CJ06789; --
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0005306; P:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 1250 AA; 147549 MW; A367A4D74887B235 CRC64;
SQ

Query Match 26.5%; Score 232; DB 2; Length 1250;
Best Local Similarity 33.7%; Fred. No. 1.5e-08;
Matches 55; Conservative 34; Mismatches 68; Indels 6; Gaps 30

QY 3 TOHILASILERSIIIFILRPIFIATIEDNDIPKILQENKLYRSSISPRYFF--YDENTKK 60
DB 260 TRADUKLIVNSLPIFFRLFIATFDK--FEILEKHKFKKSISRTILENLOEDSS 317
QY 61 KLEYKKIITIFNLDKGSDAKFFVFNFGGLFSEDKVLYNNEGLISISEIILVKMLFF 120
DB 318 SGCPGELENIFNYNKGKGNFMPVFNFGGLFDESKTALLSTPKFNDKDLKFIILNQLNF 377
QY 121 EKNIKDEKFKVYSRLDPKSGELYETLLEYDLRIADTVHRI 163
DB 378 KDKNLSFKR--DYKTLSEHLGTIVEGLLSYFFEIANEDIYVV 418

RESULT 7
Q9PPL7 CAMJE
ID Q9PPL7_CAMJE PRELIMINARY; PRT; 1250 AA.
AC Q9PPL7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Possible restriction/modification enzyme.
GN OrderedLocusNames=CJ0690c;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
NCBI_TaxID=197;
[1]
NUCLEOTIDE SEQUENCE.
RP STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204; DOI=10.1038/35001088;
RA Parkhill J., Wren B., Mungall K.L., Ketley J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moulé S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,

RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 reveals hypervariable sequences.";
 RL Nature 403:665-668 (2000).
 DR EMBL; AL139076; CAB72964.1; -; Genomic_DNA.
 DR PIR; E81339; E81339.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPR007409; DUF450.
 DR InterPro; IPR002296; N12N6_mtfase.
 DR InterPro; IPR002052; N6_Mtase.
 DR Pfam; PF04313; HSDR N; I.
 DR PRINTS; PR00507; N12N6MTFRASE.
 DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 1250 AA; 147698 MW; 0A2CD9C5A464CDEA CRC64;

 Query Match 26.5%; Score 232; DB 2; Length 1250;
 Best Local Similarity 33.7%; Pred. No. 1.5e-08;
 Matches 55; Conservative 34; Mismatches 68; Indels 6; Gaps 3;

 QY 3 TQHLASILEESIFILRIFFIAYIEDNDIFKILQENKLYRSSISFRYFF--YDENTKK 60
 Db TKADLKLIVENSIFYFRLFIAYFEDK--FEILEKHVKYFKSKISLRTLLENLQDESS 317

 QY 61 KLEYKKIITFNLLDKGSDAIFKPVFNGGLFSDKVKYLNNEGLLSISEIEEILVKMLPF 120
 Db SGGFGELENIFNYNKGKGNFDPVFNGLFSDKSTALLSTPKIFNDKOLKILNQLLNF 377

 QY 121 EKNINDEKVKYSLRDPKSGFGLYETLLDYLRDIADTTVHRI 163
 Db KKNLSFKR--DYKTLVSVEHLGIYEGLLSYFPEIANEDIYV 418

 RESULT 8
 Q9HI06 HALSA
 ID Q9HI06 HALSA PRELIMINARY; PRT; 662 AA.
 AC Q9HI06;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Vng6135c.
 GN OrderedLocusNames=VNG6135C;
 OS Halobacterium salinarum (Halobacterium halobium).
 OG Plasmid pNRC200.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=2242;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081;
 RX MEDLINE=20504483; PubMed=11016950; DOI=10.1073/pnas.190337797;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RT "Genome sequence of *Halobacterium* species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
 DR EMBL; AE005149; AAG20810.1; -; Genomic_DNA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPR002296; N12N6_mtfase.
 DR PRINTS; PR00507; N12N6MTFRASE.
 KW Complete proteome; Plasmid.
 SQ SEQUENCE 662 AA; 75448 MW; 9B5E98D6F315D65F CRC64;

 Query Match 20.9%; Score 182.5; DB 2; Length 662;
 Best Local Similarity 27.6%; Pred. No. 3.8e-05;
 Matches 55; Conservative 39; Mismatches 58; Indels 47; Gaps 10;

Best Local Similarity 27.6%; Pred. No. 3.5e-05;
 Matches 55; Conservative 39; Mismatches 58; Indels 47; Gaps 10;

 QY 10 ILEESIIIFILRIFFIAYIEDNDIFKILQ-ENKLYRSSISF-----RYF 52
 Db IHDSLLIYRLIFVLVYAEAG--RELLDTSNIEYSYSLNSLQKQETAEELDSDGPKYR 287

 QY 53 FYDENTKKLEYKIIITFNLLDKGS-----DAIKPFVFNGLFSED-----VKY 98
 Db DWQDNLQSLD-----ELFTLINKSKSGRIGPEEDLYPAYNGGLFRTDPDEDDSAEAKF 342

 QY 99 LNNE--GLLSISEIEEILVKMLFFBEKNIKDK-FVKYSRLDPKSGFGLYETLLDYDLRI 155
 Db LANHDVGDAYLAKVVELLTR----SKNNDGGGKIFVDYSSLDVRHLGSIYEGILLEYQLNV 398

 QY 156 ADTTVHRIEDGVYLIRTE 174
 Db ADEPL--ALDDGEYVSADE 415

 RESULT 9
 OS2000 HALSA
 ID OS2000 HALSA PRELIMINARY; PRT; 706 AA.
 AC OS2000;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=VNG5138C;
 OS Halobacterium salinarum (Halobacterium halobium).
 OG Plasmid pNRC100.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=2242;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081;
 RX MEDLINE=99063795; PubMed=9847077;
 RA Ng W.V., Clufo S.A., Smith T.M., Bumgarner R.E., Baskin D., Faust J.,
 RA Hall B., Loretz C., Seto J., Siagel J., Hood L., DasSarma S.;
 RT "Snapshot of a large dynamic replicon in a halophilic archaeon:
 megaplasmid or minichromosome?";
 RL Genome Res. 8:1131-1141 (1998).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081;
 RX MEDLINE=20504483; PubMed=11016950; DOI=10.1073/pnas.190337797;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
 RT "Genome sequence of *Halobacterium* species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).
 DR EMBL; AF016485; AAC82880.1; -; Genomic_DNA.
 DR PIR; T08313; T08313.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR InterPro; IPR002296; N12N6_mtfase.
 DR PRINTS; PR00507; N12N6MTFRASE.
 KW Complete proteome; Hypothetical protein; Plasmid.
 SQ SEQUENCE 706 AA; 80793 MW; 16D0E0F772A867B2 CRC64;

 Query Match 20.9%; Score 182.5; DB 2; Length 706;
 Best Local Similarity 27.6%; Pred. No. 3.8e-05;
 Matches 55; Conservative 39; Mismatches 58; Indels 47; Gaps 10;

 QY 10 ILEESIIIFILRIFFIAYIEDNDIFKILQ-ENKLYRSSISF-----RYF 52
 Db IHDSLLIYRLIFVLVYAEAG--RELLDTSNIEYSYSLNSLQKQETAEELDSDGPKYR 287

Db 274 IHDSLLIYVRLIFVLYAABG--RELLDTSNIEYQSLSNLSKQEIABELDSGPKYR 331
 QY 53 FYDENTKKLEKKIITIFNLLDKS-----DAIKFPVNGGLFSED-----KVKY 98
 Db 332 DWQDNQSLRD-----ELFTLINKSGSRGIPREDLYIPAYNGGLFRTDDEDDSAEKF 386
 QY 99 LNNE--GLASIEBIEILVMKLFEEKNHDEK-FVKYSLRDPKSPGELYETLLELDRI 155
 Db 387 LANHDVGDAYLAKVWELLTR----SKNNDGGKIFVDYSSLDVRLHSLGIEGLLEYQLNV 442
 QY 156 ADTIVRHIEDGVLRTE 174
 Db 443 ADEPL--ALDDGYVSADE 459

RESULT 10

ID Q50N37 ENTHI PRELIMINARY; PRT; 1193 AA.
 AC Q50N37;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Hypothetical protein.
 GN ORFNames=465.t00009;
 OS Entamoeba histolytica HM-1:IMSS.
 OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_TaxID=294381;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=HM-1:IMSS;
 RX PubMed=15729342; DOI=10.1038/nature03291;
 RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
 RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
 RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
 RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
 RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
 RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
 RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
 RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
 RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
 RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
 RA Fraser C.M., Hall N.;
 RT "The genome of the protist parasite Entamoeba histolytica."
 RL Nature 433:865-868(2005).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAFB01001119; EAL43006.1; -: Genomic_DNA.
 KW Hypothetical protein.
 SQ SEQUENCE 1193 AA; 138539 MW; A3E2153B243C031B CRC64;

Query Match 12.5%; Score 109.5; DB 2; Length 1193;
 Best Local Similarity 27.9%; Pred. No. 15;
 Matches 50; Conservative 30; Mismatches 64; Indels 35; Gaps 11;
 QY 28 EDNDIFKKILQENKLYRSSISFRYF---FYDENTKKLE--YKKII-----TIFNLLDKG 77
 Db 58 EMNVIVKKISEENVFYMRSEYETFLKFTTNMKLSDOQFYKFNKIKVFNFKIKE 117
 QY 78 SD-----AIKFPVNGGLFSEDVKVYLNNEGLLSISEI---EEI-----LVKMLFFEE 122
 Db 118 TWYTELILKFTFYLNLF--SRIELVDEGKLFGLDFNVEEFKSKTVNIEKAKYIEE 176
 QY 123 KNKDEKFKVYSLRDPKSPGELYETLLELDRIAD---TTVHRIL-EDGVYLI---RTE 174
 Db 177 VPLVVEFIIIEINKID--GFNKIYEMIRLYDLKLLNSMLTSVSYFSDNNMYLLIEDQORTE 233

RESULT 11

Q510L7 ENTHI PRELIMINARY; PRT; 2181 AA.
 ID Q510L7 ENTHI PRELIMINARY;
 AC Q510L7;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 GN Ubiquitin carboxyl-terminal hydrolase, putative.
 GN ORFNames=109.t00017;
 OS Entamoeba histolytica HM-1:IMSS.
 OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_TaxID=294381;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=HM-1:IMSS;
 RX PubMed=15729342; DOI=10.1038/nature03291;
 RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
 RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
 RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
 RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
 RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
 RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
 RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
 RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
 RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
 RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
 RA Fraser C.M., Hall N.;
 RT "The genome of the protist parasite Entamoeba histolytica."
 RL Nature 433:865-868(2005).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAFB01000388; EAL47296.1; -: Genomic_DNA.
 KW Hydrolase.
 SQ SEQUENCE 2181 AA; 253911 MW; 4968693964C71AAE CRC64;
 Query Match 12.5%; Score 109.5; DB 2; Length 2181;
 Best Local Similarity 27.9%; Pred. No. 28;
 Matches 50; Conservative 30; Mismatches 64; Indels 35; Gaps 11;
 QY 28 EDNDIFKKILQENKLYRSSISFRYF---FYDENTKKLE--YKKII-----TIFNLLDKG 77
 Db 58 EMNVIVKKISEENVFYMRSEYETFLKFTTNMKLSDOQFYKFNKIKVFNFKIKE 117
 QY 78 SD-----AIKFPVNGGLFSEDVKVYLNNEGLLSISEI---EEI-----LVKMLFFEE 122
 Db 118 TWYTELILKFTFYLNLF--SRIELVDEGKLFGLDFNVEEFKSKTVNIEKAKYIEE 176
 QY 123 KNKDEKFKVYSLRDPKSPGELYETLLELDRIAD---TTVHRIL-EDGVYLI---RTE 174
 Db 177 VPLVVEFIIIEINKID--GFNKIYEMIRLYDLKLLNSMLTSVSYFSDNNMYLLIEDQORTE 233
 RESULT 12
 Q4XXL0 PLACH PRELIMINARY; PRT; 2337 AA.
 ID Q4XXL0 PLACH PRELIMINARY;
 AC Q4XXL0;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN ORFNames=PC001007.02.0;
 OS Plasmodium chabaudi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5825;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
 RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
 RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
 RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
 RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
 RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
 RT "A comprehensive survey of the Plasmodium life cycle by genomic,
 RT transcriptomic, and proteomic analyses."
 RL Science 307:82-86(2005).
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

```

CC preliminary data.
DR EMBL: CAJ01002429; CAH78351.1; -; Genomic_DNA.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 2337 AA; 289397 MW; 0DCA834991D6A8CA CRC64;

Query Match 12.4%; Score 108.5; DB 2; Length 2337;
Best Local Similarity 27.2%; Pred. No. 36;
Matches 41; Conservative 26; Mismatches 55; Indels 29; Gaps 6;

QY 5 HILALISERSIFILRIFFIATYEDNDIFKKILQENKLYRSSI-SPRYFFYDENTKKLE 63
Db 1948 HNWVSIVEKLIYVNI-----NEI-KKILKKNKIFQVIRFYFNFKLKDDMLQ 1998

QY 64 YKKIITFNLIDKGSDAIKFPVFNGLFSEDKVKYL-----NNEGLLSI 107
Db 1999 KKYFQIVYNLVKKNISINKLQIVYSSLOVKRMKYLFFYNRRKCNTRKKLQNYTHLS- 2057

QY 108 SEIEEILV--KMLFFFEKNIKDKFKVYSRL 136
Db 2058 KKKHSIFTRKWAIFIRKNINDEKLAFNKL 2088

RESULT 13
Q54H63_DICDI
ID Q54H63_DICDI PRELIMINARY; PRT; 868 AA.
AC Q54H63;
DT 13-SRP-2005 (TrEMBLrel. 31, Created)
DT 13-SRP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SRP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=DD80188518;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Safranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Louissege H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Plattner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAFL01000219; EAL62554.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 868 AA; 105066 MW; D7DCD2A3EE249CC1 CRC64;

Query Match 12.3%; Score 107.5; DB 2; Length 868;
Best Local Similarity 24.7%; Pred. No. 15;
Matches 40; Conservative 36; Mismatches 55; Indels 31; Gaps 7;

QY 19 LRIFFIAYIE--DNDIF-----KKILQENKLYRSSIS--PRYFFYDENTKKLYK--- 65
Db 284 LKEFFKSYIKSINNNSPKFKENKDIETNNLVKKYIKKFKLVLLVDYENGKKKKDKFKIPD 343

```

```

QY 66 -KII-----TIFNLIDKGSDAIKFPVFNGLFSEDKVKYLNNEGLLSISEIEE 112
Db 344 LKEIESMKFTNQFNSTISDLITDGDDE---NYEKDI---QENLRILYSLAPFFIEQYDH 398

QY 113 ILVKMLFFFEKNIKDKFKVYSRLDPKSGFELYETLLLEYDLR 154
Db 399 LIGNLIFYKVKILNEKDFMKYNHLYPKLCPKMGIVGIYDYK 440

RESULT 14
Q5FM50_LACAC
ID Q5FM50_LACAC PRELIMINARY; PRT; 939 AA.
AC Q5FM50;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Methylase.
GN OrderedLocusNames=LBA0332;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCFM;
RA Altermann E., Russell W.M., Azcarate-Perill M.A., Barrangou R.,
RA Buck B.L., McAuliffe O., Southern N., Dobson A., Duong T., Callanan M.,
RA Lick S., Hamrick A., Cano R., Klammer T.R.;
RT "Complete genome sequence of the probiotic lactic acid bacterium
RT Lactobacillus acidophilus NCFM.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:3906-3912(2005).
DR EMBL: CP000033; AA42224.1; -; Genomic_DNA.
DR GO: GO:003677; F:DNA binding; IEA.
DR GO: GO:0008170; F:N-methyltransferase activity; IEA.
DR GO: GO:0006306; P:DNA methylation; IEA.
DR InterPro: IPR002296; N12N6_mtfase.
DR PRINTS: PR0507; N12N6MTRASE.
DR KW Complete proteome; Methyltransferase.
DR SEQUENCE 939 AA; 108344 MW; 4635DB98E12A08C CRC64;

Query Match 12.3%; Score 107.5; DB 2; Length 939;
Best Local Similarity 25.0%; Pred. No. 17;
Matches 42; Conservative 25; Mismatches 52; Indels 49; Gaps 8;

QY 19 LRIFFIAYIEDNDIFKKILQENKLYRSSISFRYFFYD--ENTKKLYKKIIFNLLD- 75
Db 208 VRIVFCLYAEDAGIFGK-----KNMFHDYLEEFDARHMRKALINLFKVLDT 253

QY 76 --KGSD-----AIKFPVFNGLFSEDKVKYLNNEGLLSISEIEEILVKMLFPEKN 124
Db 254 KVXDRDPYLEDDSPKLAQFPYVNGMFSDE-----NIEIPPTDELKDLLL-----SK 301

QY 125 IKDEKFKVYSRLDPKSGFELYETLLLEYDLR-----IADTVHRIIE 165
Db 302 ASDS--FDWSEISPTIFGAVFESTLNPDRRQGMHYTSIENHKVID 347

RESULT 15
Q8FSZ4_COREF
ID Q8FSZ4_COREF PRELIMINARY; PRT; 1524 AA.
AC Q8FSZ4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=CE2898;
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Y8-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;
RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RT "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of *Corynebacterium*
RT efficiens";
RL Genome Res. 13:1572-1579(2003).
DR ENBL; BA000035; BAC19708.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1524 AA; 171619 MW; 6FCE5647350F256E CRC64;

Query Match 12.2%; Score 106.5; DB 2; Length 1524;
Best Local Similarity 20.8%; Pred. No. 33;
Matches 42; Conservative 37; Mismatches 58; Indels 65; Gaps 8;

QY 10 ILRESIIFILRIFPIAYIE---DNDIFPKILQENKLYRSSISFRYPFYE----- 56
Db 281 LAKQSLRYLRIILFLFAEASPELEILFTGTPE-----YDEGYGLSLREL 326

QY 57 -----NTKKLE-YKKIITIFNLIDKGD-----AIKFPVFNGLF 91
Db 327 ILTPPVTONARSGTHLYDSQLFLHLVDQNDPRDSNPDPDPAGEGLQFRNLSADLF 386

QY 92 SE-----DKVYNNEGLLSISEIEILVKMLFFEEKNIKDEKFKYKSLDPKPSFGELY 145
Db 387 QPKATNIIDTVK-LSNEAL-----NTVLENLLLSKEQSGKRGFISYATLGVTELQVY 439

QY 146 ETLLEYDLRIADTVVHRIIEDG 167
Db 440 EGLMSYTGFIAREEDLFEVAKNG 461

Search completed: January 24, 2006, 19:56:50
Job time : 39.3889 secs

This Page Blank (uspto)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	99	11.3	661	2	US-09-487-558B-166	Sequence 166, App
2	97	11.1	373	1	US-08-015-986A-14	Sequence 14, Appl
3	97	11.1	373	1	US-08-446-363-14	Sequence 14, Appl
4	97	11.1	2308	1	US-08-015-973-1	Sequence 1, Appl
5	97	11.1	2308	1	US-08-448-164-1	Sequence 1, Appl
6	97	11.1	2308	2	US-08-081-929-2	Sequence 2, Appl
7	97	11.1	2308	2	US-10-000-954-2	Sequence 2, Appl
8	97	11.1	2314	2	US-09-816-703A-2	Sequence 2, Appl
9	96	11.0	404	2	US-09-134-001C-4848	Sequence 4848, Ap
10	95.5	10.9	331	2	US-09-710-279-3256	Sequence 3256, Ap
11	95.5	10.9	585	2	US-09-107-532A-5494	Sequence 5494, Ap
12	93.5	10.7	622	2	US-09-605-703B-604	Sequence 604, App
13	93.5	10.7	1218	2	US-09-605-703B-600	Sequence 600, App
14	93	10.6	389	2	US-09-800-170-18	Sequence 18, Appl
15	93	10.6	408	2	US-09-107-532A-6196	Sequence 6196, Ap
16	90	10.3	482	2	US-09-710-279-24	Sequence 24, Appl
17	90	10.3	574	2	US-09-134-001C-5409	Sequence 5409, Ap
18	89.5	10.2	249	2	US-09-710-279-3298	Sequence 3298, Ap
19	88	10.1	287	2	US-09-107-532A-4644	Sequence 4644, Ap
20	86	9.8	335	2	US-09-134-000C-4305	Sequence 4305, Ap
21	86	9.8	511	2	US-09-107-532A-6171	Sequence 6171, Ap
22	86	9.8	625	2	US-09-583-545-2	Sequence 2, Appl
23	85	9.8	625	2	US-09-617-594A-6	Sequence 6, Appl
24	86	9.8	625	2	US-10-209-507-6	Sequence 6, Appl
25	86	9.8	870	2	US-09-172-339-8	Sequence 8, Appl
26	85.5	9.8	319	2	US-09-107-532A-4723	Sequence 4723, Ap
27	85.5	9.8	341	2	US-09-248-796A-36889	Sequence 26889, A

```
US-08-015-986A-14
; Sequence 14, Application US/08015986A
; Patent No. 5532123
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-GAMMA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,986A
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-015-986A-14

Query Match 11.1%; Score 97; DB 1; Length 373;
Best Local Similarity 21.8%; Pred. No. 0.063;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNDIF----KKILOENKLYRSISFRYPFFYDENTKKKLEYKKII-----68
Db 123 YCFDADRFSSFEAAVKGKGLRALSLF-----EVGTEENLDKFAIIDGVESVSRFGKQA 177
Qy 69 -----TIFNLDKGSDAIKFPVNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
Db 178 ALDPPILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 235
Qy 115 -----VKMLFFEEKNIDKFKVYSR-----LDPKSFGE 143
Db 236 MQQSGYVWMLDYLNQNFREQY-KFSRQVFSYTGKEIHEAVCSSEPENVAQDPENYTS 294
Qy 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 295 LLVTWERPRVVYDTMIKFAVLYQQLDGEDQTKHEFLTDG 334

RESULT 3
US-08-446-363-14
; Sequence 14, Application US/08446363
; Patent No. 5891700
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-GAMMA
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,986A
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 373 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-015-986A-14

Query Match 11.1%; Score 97; DB 1; Length 373;
Best Local Similarity 21.8%; Pred. No. 0.063;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNDIF----KKILOENKLYRSISFRYPFFYDENTKKKLEYKKII-----68
Db 123 YCFDADRFSSFEAAVKGKGLRALSLF-----EVGTEENLDKFAIIDGVESVSRFGKQA 177
Qy 69 -----TIFNLDKGSDAIKFPVNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
Db 178 ALDPPILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 235
Qy 115 -----VKMLFFEEKNIDKFKVYSR-----LDPKSFGE 143
Db 236 MQQSGYVWMLDYLNQNFREQY-KFSRQVFSYTGKEIHEAVCSSEPENVAQDPENYTS 294
Qy 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 295 LLVTWERPRVVYDTMIKFAVLYQQLDGEDQTKHEFLTDG 334

RESULT 4
US-08-015-973-1
; Sequence 1, Application US/08015973
; Patent No. 5604094
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASE-BETA
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; US-08-446-363-14
```


[illegible]

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2308 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-081-929-2

Query Match      11.1%; Score 97; DB 2; Length 2308;
Best Local Similarity 21.8%; Pred. No. 0.76;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKILOENKLYRSSISFRYFFYDENTKKKLEYKKII-----68
Db 156 YCFDADRPFSSFEAEVKGKGLRALSILF-----EVGTEENLDKAIIDGVESVSRFGKQA 210
QY 69 -----TIFNLLDKGSDAIKFPVNGGLFS---EDVKYLNNEGLLSISRIE-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPPTDVTVDWIVFKDVTVSISESQLAVFCEVLT 268
QY 115 -----VKMLFFEEKNIKDKFKVYSR-----LDPKSFGE 143
Db 269 MQOSGYVLMMDYLQNNFREQQY-KFSRQVFSSTYTGKEIHEAVCSSEPENVQADPENYTS 327
QY 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
Db 328 LLVTWERPRVVDYTMIEKFAVLVYQQLDGEDQTKHEFLTDG 367

RESULT 7
US-10-000-954-2
; Sequence 2, Application US/10000954
; Patent No. 6803452
; GENERAL INFORMATION:
; APPLICANT: Schlusseringer, Joseph
; Barnea, Gilad
; Grumet, Martin H.
; Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RPTASES: THEIR
; STRUCTURAL DOMAINS AND LIGANDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/000,954
; FILING DATE: 04-Dec-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/644,293
; FILING DATE: 23-Aug-2000
; APPLICATION NUMBER: 08/081,929
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2308 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-000-954-2

Query Match      11.1%; Score 97; DB 2; Length 2308;
Best Local Similarity 21.8%; Pred. No. 0.76;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKILOENKLYRSSISFRYFFYDENTKKKLEYKKII-----68
Db 156 YCFDADRPFSSFEAEVKGKGLRALSILF-----EVGTEENLDKAIIDGVESVSRFGKQA 210
QY 69 -----TIFNLLDKGSDAIKFPVNGGLFS---EDVKYLNNEGLLSISRIE-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPPTDVTVDWIVFKDVTVSISESQLAVFCEVLT 268
QY 115 -----VKMLFFEEKNIKDKFKVYSR-----LDPKSFGE 143
Db 269 MQOSGYVLMMDYLQNNFREQQY-KFSRQVFSSTYTGKEIHEAVCSSEPENVQADPENYTS 327
QY 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
Db 328 LLVTWERPRVVDYTMIEKFAVLVYQQLDGEDQTKHEFLTDG 367

RESULT 8
US-09-816-703A-2
; Sequence 2, Application US/09816703A
; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 AGY
; CURRENT APPLICATION NUMBER: US/09/816,703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992)
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Levy, J.B., et al.: The cloning of a receptor-type protein tyrosi
; OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
; OTHER INFORMATION: Chem. 268 (14), 10573-10581 (1993)
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION: By similarity
; NAME/KEY: DOMAIN
; LOCATION: (25)..(1635)
; OTHER INFORMATION: Extracellular (potential)
; NAME/KEY: misc_feature
; LOCATION: (25)..(2314)
; OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
; NAME/KEY: DOMAIN
; LOCATION: (34)..(302)
; OTHER INFORMATION: Carbonic-anhydrase like
; NAME/KEY: SITE
; LOCATION: (105)..(105)
```

```
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (134)..(134)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (223)..(223)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: DOMAIN
; LOCATION: (312)..(406)
; OTHER INFORMATION: Fibronectin Type-III
; NAME/KEY: SITE
; LOCATION: (324)..(324)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (381)..(381)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (497)..(497)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (552)..(552)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (587)..(587)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (602)..(602)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (629)..(629)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (637)..(637)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (677)..(677)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: VARSPLIC
; LOCATION: (755)..(1614)
; OTHER INFORMATION: Splicing variant; missing (in short isoform)
; NAME/KEY: BINDING
; LOCATION: (997)..(997)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1017)..(1017)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1050)..(1050)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
```

```
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
; US-09-816-703A-2
```

```
Query Match 11.1%; Score 97; DB 2; Length 2314;
Best Local Similarity 21.8%; Pred. No. 0.7%;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNIF---KKIIQENKLYRSSISFRFFVDENTKKKLEYKKII-----68
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156 YCFDADRFSFEAEVKGKRLSILF-----EVGTEENLDPKAIDGVESVSRFGQA 210
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 -----TIFNLLDKGSDAIFPFVENGGLFS---BDKVYLNNEGLISTSEIE---BIL- 114
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCSEVLT 269
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 115 -----VKMLFFPEEKIKDEKFKVYSR-----LDPKSFGE 143
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 269 MQQSGYVWMLDYLNQNNFREQY-KFSRQVFSSYTGKEEIHAEVCSSEPVQADPENYTS 327
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 144 L-----YETILE-----YDLRIADTTVHRIEDG 167
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 LLVTWRPRVYDTMIEKFAVLYQQLDGEDQTKHEFLTDG 367
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 9
US-09-134-001C-4848
; Sequence 4848, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4848
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-4848
```

```
Query Match 11.0%; Score 96; DB 2; Length 404;
Best Local Similarity 26.4%; Pred. No. 0.089;
Matches 43; Conservative 25; Mismatches 51; Indels 44; Gaps 9;
```


Query Match 10.7%; Score 93.5; DB 2; Length 622;
Best Local Similarity 22.5%; Pred. No. 0.3;
Matches 27; Conservative 22; Mismatches 54; Indels 17; Gaps 2;

Qy 64 YKKIITIFNLDDK-----GSDAIKFPVFGGLPSEDKVKYLNNEGLLSI 107
Db 6 YDSLQLLFTLVDKGHHPTDAKAVAFDAEAGEGLHFRNLSADLFLPAATELIDRVG-LSN 64
Qy 108 SEIBEILVKMLFFBEKNIKDEKFKYKSRLOPKSGELYETLLLEYDLRIADTTVHRIIEDG 167
Db 65 EALNKNVLELLSRVQSGKDRGFISYATLGTVELGQVYEGLSMTGTGFIQBEDLPEVAPHG 124

RESULT 13

US-09-605-703B-600
; Sequence 600, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oekar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE OF INVENTION: PROTEINS
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 600
; LENGTH: 1218
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-600

Query Match 10.7%; Score 93.5; DB 2; Length 1218;
Best Local Similarity 22.5%; Pred. No. 0.74;
Matches 27; Conservative 22; Mismatches 54; Indels 17; Gaps 2;

Qy 64 YKKIITIFNLDDK-----GSDAIKFPVFGGLPSEDKVKYLNNEGLLSI 107
Db 27 YDSLQLLFTLVDKGHHPTDAKAVAFDAEAGEGLHFRNLSADLFLPAATELIDRVG-LSN 85
Qy 108 SEIBEILVKMLFFBEKNIKDEKFKYKSRLOPKSGELYETLLLEYDLRIADTTVHRIIEDG 167
Db 86 EALNKNVLELLSRVQSGKDRGFISYATLGTVELGQVYEGLSMTGTGFIQBEDLPEVAPHG 145

RESULT 14

US-09-800-170-18
; Sequence 18, Application US/09800170
; Patent No. 6481667
; GENERAL INFORMATION:
; APPLICANT: Kinsella, Todd
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES
; FILE REFERENCE: A-68614-1/DJB/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/800,170
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-800-170-18

Query Match 10.6%; Score 93; DB 2; Length 389;
Best Local Similarity 19.5%; Pred. No. 0.18;
Matches 43; Conservative 41; Mismatches 76; Indels 60; Gaps 7;

Qy 5 HILASILEESIIIFILRIFFIAYIEDN-----DIFKK----- 35
Db 179 YIEVPLFEDPLLLLPFGIVSRLEKSTKLISGKRNLELFRKHVGFDTSEKAKALDELISK 238
Qy 36 -----ILQENKLYRSSISPRYFFYDENTKKLEYKKIITIFNLDDKGS--AIKFPV 85
Db 239 AKESERYPILEE--LRLGLLFGFTRNELRIENPTVVLMEILERIEGSPNLAEKIIV 296
Qy 86 FNGCLPSEDKVKYLNNEGLLSISEIBEILVKMLFFBEKNIKDEKFKYKSRLOPKSGELY 145
Db 297 LEGRIKEENYLRILEEGLIENGKLTGELKELLEVRNRREFSKDQVDYVR-----NIV 349
Qy 146 ETLL-----LEYDLRIADTT--VHRIEDGVVLI RT 173
Db 350 ENLVFLPVEKVERIEYGVYDVTTHFVANGILVHNT 389

RESULT 15

US-09-107-532A-6196
; Sequence 6196, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6196:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...408
; SEQUENCE DESCRIPTION: SEQ ID NO: 6196:
US-09-107-532A-6196

Query Match 10.6%; Score 93; DB 2; Length 408;

Best Local Similarity 22.6%; Pred. No. 0.19;
 Matches 51; Conservative 30; Mismatches 47; Indels 98; Gaps 12;

Qy	10	ILBESIIFILRIPIFIAYIEDNDIFKKILOENKLYRSSISFRFYDYDENTKKLEYKKIIT	69
Db	14	VLLKSLHVCCLKPIWVFNSEFSIKE-----YRSDRT-TYFFYD-----YKGLLT	57
Qy	70	IFNLLDKGSDAIK-----FPVFNGGLFSE-----DKVKY-----	98
Db	58	-----DNIKKTKDFHFTGD-FNEMFLLYMFKTRYVYLFGPFRANNIDKVFFKLKMN	107
Qy	99	-----LNNEGLLSISEIEIILVKMLPFEKNIKD---EKFVKYS-----	134
Db	108	NLNVAMADRERLYNSLQNLTLYSLGDIRDILILVHYFTGKIEDLFHEPLIEYTENLSKT	167
Qy	135	-----RLDPKSGELYET-LLEY-----DLRIADTTVHRI	163
Db	168	IEQIKIDNLLSQNYDPEIYLFYENKILEYVKNGDIRNLENMVFNL	213

Search completed: January 24, 2006, 20:02:43
 Job time : 11.7267 secs


```

Qy 16 IFILRFPAYIEDNDIFKILQENKLYRSIS-----FRYFFYDENTKKLEYKKI 67
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 188 LHVRLFLCLFADTTIFKSLFYQYIETKLDGSLAHINTFLVLTNPQSKLKN- 246
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 68 ITIFNLLDKGSDAIFPVPVNGLFSE-----DKVKYLNNEGILLSISEIILVKMLFF 120
      ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 247 -----LDEHLAA--FPYINGKLFEPLPAQPOKAM--REALDLCSLD- 286
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 121 EEKNIKDERFVKYSLRDPKSGFELYETILLE 150
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 287 -----WSRISPAIFGSLFOSIMD 304
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 6

```

US-10-437-963-187688
; Sequence 187688, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187688
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(392)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84367C.1.pep
US-10-437-963-187688

```

Query Match	11.3%;	Score 99;	DB 4;	Length 392;
Best Local Similarity	28.3%;	Pred. No. 1.1;		
Matches	34;	Conservative 27;	Mismatches 41;	Indels 18; Gaps 5;

Qy	10	I L E S I I F I L R I F F A Y I E N D I P K I L Q E N K L Y R S I S F Y F Y D E N T K K L Y K I --	67
	:	: :	
	:	: :	
Db	83	V L I N Y V L S S L W M F I S F L --- I P K G V L Q N I D Y Y R S --- R F F W K D D H K K K K Y R L R W	134
Qy	68	- I I F N L L D K G S D A I - K P F V N F G L F S E D K V K Y I N N E G L L S I S I E I L L V P O L F P E K N I	125
	:	: :	
	:	: ~	
Db	135	D I I C Q K X I D O G F G V H N L E V N Q N K L L S K X I L F K I N E O G ----- V W K D L L K M K F N Y C O I	188

RESULT. T 7

US-09-801-368-166
; Application 166, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie

```
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53613
; LENGTH: 1032
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-282-122A-53613

Query Match      11.1%; Score 99; DB 4; Length 1032;
Best Local Similarity 27.5%; Pred. No. 3.6; Mismatches 28; Indels 38; Gaps 7;
Matches 44; Conservative 28;

Qy 34 KKILQENKLYRSSISFRYPFYDENTK---KKLEYKKIIT--IFNLLDKG--SDAIKFPVF 86
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 88 KLIKWEGKHARLEIALDITEKEDISIAEKLEIKELLVGCIKHLITESNFSDAVNLVLT 147
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 87 NGGLF-----SEDKVYLN-----NEGL-----LSISEIEIILVKMLFPFEK- 123
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 NIGKFKRADRAYFVSEDRKVTNTYWCNCGYDEQKQALTNLDINELKNWFLFEQKG 207
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 124 -----NIKDKFKVYSRLDPKSGFGLYETLLLEYDLRI 155
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 FVIIDDIENIKEQPIEYILKPKIESLMTSSLKQDDKV 247
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-983-000A-2
; Sequence 2, Application US/09983000A
; Publication No. US20030118585A1
; GENERAL INFORMATION:
; APPLICANT: AGY Therapeutics
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: USE OF PROTEIN BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALI
; FILE REFERENCE: 263/180 -- Peagleman -- AGY
; CURRENT APPLICATION NUMBER: US/09/983,000A
; CURRENT FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: gene
; LOCATION: (1)..(374)
; OTHER INFORMATION: PTP-zeta SM1 exon 9 variant
US-09-983-000A-2

Query Match      11.1%; Score 97; DB 3; Length 374;
Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNDIF----KKILOENKLYRSSISFRYPFYDENTKKLEYKKII----- 68
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 YCFDADRFSSFEAVKGGKLRALSILF-----EVGTENLDFKALIDGVESVSRFGKQA 210
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 -----TIFNLLDKGSDAIKFPVFNGLFS---EDKVYLNNEGLLSISEIE-----BIL- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDVTVDWIVFKDVTVSISQSLAVFCEVLT 268
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 115 -----VKMLFPEEKNIKDEKFKVYSR-----LDPKSFGE 143
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 MQQSGYVMLDYQLNNFREQY--KFSRQVFSYTGKEEIHAVCSSEPENVQADPENYTS 327
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 LLVWTERPRVVDYTMIEKFAVLVYQLDGEDQTKHEFLTGD 367
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-10-723-860-951
; Sequence 951, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05862.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 951
; LENGTH: 1005
; TYPE: PRT
```

```
Db 328 LLVWTERPRVVDYTMIEKFAVLVYQLDGEDQTKHEFLTGD 367
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-10-652-981-2
; Sequence 2, Application US/10652981
; Publication No. US20050074400A1
; GENERAL INFORMATION:
; APPLICANT: AGY Therapeutics
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: USE OF PROTEIN BIOMOLECULAR TARGETS IN THE TREATMENT AND VISUALI
; FILE REFERENCE: 263/180 -- Peagleman -- AGY
; CURRENT APPLICATION NUMBER: US/10/652,981
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: US/09/983,000A
; PRIOR FILING DATE: 2001-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: gene
; LOCATION: (1)..(374)
; OTHER INFORMATION: PTP-zeta SM1 exon 9 variant
US-10-652-981-2

Query Match      11.1%; Score 97; DB 5; Length 374;
Best Local Similarity 21.8%; Pred. No. 1.6;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNDIF----KKILOENKLYRSSISFRYPFYDENTKKLEYKKII----- 68
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 YCFDADRFSSFEAVKGGKLRALSILF-----EVGTENLDFKALIDGVESVSRFGKQA 210
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 -----TIFNLLDKGSDAIKFPVFNGLFS---EDKVYLNNEGLLSISEIE-----BIL- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDVTVDWIVFKDVTVSISQSLAVFCEVLT 268
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 115 -----VKMLFPEEKNIKDEKFKVYSR-----LDPKSFGE 143
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 269 MQQSGYVMLDYQLNNFREQY--KFSRQVFSYTGKEEIHAVCSSEPENVQADPENYTS 327
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 LLVWTERPRVVDYTMIEKFAVLVYQLDGEDQTKHEFLTGD 367
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
US-10-723-860-951
; Sequence 951, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05862.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 951
; LENGTH: 1005
; TYPE: PRT
```

; ORGANISM: Homo sapiens
US-10-723-860-951

Query Match 11.1%; Score 97; DB 5; Length 1005;
Best Local Similarity 21.8%; Pred. No. 5.3;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKILQENKLYRSSISFRFFVDENTKKLEYKKII-----68
Db 156 YCFDADRFSFEAVKGGKLRALSILF-----EVGTEENLDFKAIIDGVESVSFRGKQA 210

QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDVTVSISESQLAVFCEVLT 268

QY 115 -----VKMLFFPEKNIKDEKFKVYSR-----LDPKSFGE 143
Db 269 MQQSGYVLMYLDLQNNFREQQY--KFSRQVFSYTGKEEIIHEAACRSEPNVQADPENYTS 327

QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 328 LLVTWPRVVDYTMIEKFAVLVYQQLDGEDQTKHEFLTGD 367

RESULT 12

US-10-732-923-19145
; Sequence 19145, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 19145
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-732-923-19145

Query Match 11.1%; Score 97; DB 5; Length 1005;
Best Local Similarity 21.8%; Pred. No. 5.3;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKILQENKLYRSSISFRFFVDENTKKLEYKKII-----68
Db 156 YCFDADRFSFEAVKGGKLRALSILF-----EVGTEENLDFKAIIDGVESVSFRGKQA 210

QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDVTVSISESQLAVFCEVLT 268

QY 115 -----VKMLFFPEKNIKDEKFKVYSR-----LDPKSFGE 143
Db 269 MQQSGYVLMYLDLQNNFREQQY--KFSRQVFSYTGKEEIIHEAACRSEPNVQADPENYTS 327

QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 328 LLVTWPRVVDYTMIEKFAVLVYQQLDGEDQTKHEFLTGD 367

RESULT 13

US-10-295-027-1264
; Sequence 1264, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.

; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Ros Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1264
; LENGTH: 1113
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1264

Query Match 11.1%; Score 97; DB 4; Length 1113;
Best Local Similarity 21.8%; Pred. No. 6.1;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKILQENKLYRSSISFRFFVDENTKKLEYKKII-----68
Db 156 YCFDADRFSFEAVKGGKLRALSILF-----EVGTEENLDFKAIIDGVESVSFRGKQA 210

QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
Db 211 ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDVTVSISESQLAVFCEVLT 268

QY 115 -----VKMLFFPEKNIKDEKFKVYSR-----LDPKSFGE 143
Db 269 MQQSGYVLMYLDLQNNFREQQY--KFSRQVFSYTGKEEIIHEAACRSEPNVQADPENYTS 327

QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
Db 328 LLVTWPRVVDYTMIEKFAVLVYQQLDGEDQTKHEFLTGD 367

RESULT 14

US-10-295-027-378
; Sequence 378, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.

```
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 378
; LENGTH: 1337
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-378
```

```
Query Match 11.1%; Score 97; DB 4; Length 1337;
Best Local Similarity 21.8%; Pred. No. 7.6;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNIP-----KKILOENKLYRSSISFRYPFYDENTKKKLEVKKII----- 68
Db 38 YCFDADRFSSFEAVKGGKLRALSILF-----EVGTEENLDPKALIDGVESVSRFGKQA 92
Qy 69 -----TIFNLLDKGSDAIKFPVFNGLFS---EDKVKYLNNEGLLSISEIE-----EIL- 114
Db 93 ALDPPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 150
Qy 115 -----VKMLFFEEKNIKDEKFKVYSR-----LDPKSFGE 143
Db 151 MQQSGYVLMMDYLQNNFRQQY-KFSRQVFSSTGKEIHEAVCSSEPNVQADPENYTS 209
Qy 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
Db 210 LLVTWRPRVVDYDTMIEKFAVLQQLDGEDQTKHEFLTDG 249
```

RESULT 15

```
US-10-295-027-1266
; Sequence 1266, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glyne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
```

```
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1266
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1266
```

```
Query Match 11.1%; Score 97; DB 4; Length 1359;
Best Local Similarity 21.8%; Pred. No. 7.8;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

Qy 26 YIEDNIP-----KKILOENKLYRSSISFRYPFYDENTKKKLEVKKII----- 68
Db 156 YCFDADRFSSFEAVKGGKLRALSILF-----EVGTEENLDPKALIDGVESVSRFGKQA 210
Qy 69 -----TIFNLLDKGSDAIKFPVFNGLFS---EDKVKYLNNEGLLSISEIE-----EIL- 114
Db 211 ALDPPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESQLAVFCEVLT 268
Qy 115 -----VKMLFFEEKNIKDEKFKVYSR-----LDPKSFGE 143
Db 269 MQQSGYVLMMDYLQNNFRQQY-KFSRQVFSSTGKEIHEAVCSSEPNVQADPENYTS 327
Qy 144 L-----YETLLE-----YDLRIADTTVHRIIEDG 167
Db 328 LLVTWRPRVVDYDTMIEKFAVLQQLDGEDQTKHEFLTDG 367
```

Search completed: January 24, 2006, 20:54:33
Job time : 31.5815 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 2.78483 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-14
Perfect score: 875
Sequence: 1 QYQHILASILESIIFILR.....IADTTVHRIEDGVILIRTE 174

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA New.*
- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB_PEP.*
 - 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB_PEP.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB_PEP.*
 - 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB_PEP.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB_PEP.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB_PEP.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB_PEP.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	11.1	374	US-11-097-728-4	Sequence 4, Appli
2	97	11.1	2314	US-11-097-728-2	Sequence 2, Appli
3	97	11.1	2353	US-11-097-728-6	Sequence 6, Appli
4	95.5	10.9	331	US-10-793-626-3256	Sequence 3256, Ap
5	93	10.6	389	US-11-069-642-18	Sequence 18, Appl
6	90	10.3	482	US-10-793-626-24	Sequence 24, Appl
7	89.5	10.2	249	US-10-793-626-3298	Sequence 3298, Ap
8	82.5	9.4	2710	US-11-051-453-41	Sequence 41, Appl
9	82	9.4	833	US-10-793-626-3876	Sequence 3876, Ap
10	80	9.1	222	US-10-793-626-128	Sequence 128, App
11	80	9.1	347	US-10-793-626-106	Sequence 106, App
12	80	9.1	989	US-10-793-626-2594	Sequence 2594, Ap
13	79.5	9.1	371	US-10-793-626-3294	Sequence 3294, Ap
14	77	8.8	199	US-10-793-626-1722	Sequence 1722, Ap
15	77	8.8	312	US-11-156-084-331	Sequence 331, App
16	76.5	8.7	396	US-11-022-562-228	Sequence 228, App
17	76	8.7	398	US-11-129-143-74	Sequence 74, Appl
18	74.5	8.5	449	US-10-485-517-272	Sequence 272, App
19	74	8.5	393	US-10-485-517-316	Sequence 316, App
20	74	8.5	437	US-11-069-642-92	Sequence 92, Appl
21	74	8.5	438	US-11-069-642-47	Sequence 47, Appl
22	74	8.5	437	US-11-069-642-49	Sequence 49, Appl
23	74	8.5	438	US-11-069-642-53	Sequence 53, Appl
24	74	8.5	438	US-11-069-642-55	Sequence 55, Appl
25	74	8.5	438	US-11-069-642-57	Sequence 57, Appl

26	74	8.5	438	7	US-11-069-642-59	Sequence 59, Appl
27	74	8.5	438	7	US-11-069-642-61	Sequence 61, Appl
28	73.5	8.4	2367	7	US-11-051-453-42	Sequence 42, Appl
29	73	8.3	438	7	US-11-069-642-51	Sequence 51, Appl
30	73	8.3	438	7	US-11-069-642-63	Sequence 63, Appl
31	73	8.3	701	6	US-10-793-626-710	Sequence 710, App
32	72	8.2	413	7	US-11-032-951-12	Sequence 12, Appl
33	71.5	8.2	600	6	US-10-606-302-3	Sequence 3, Appli
34	71	8.1	289	6	US-10-793-626-880	Sequence 880, App
35	71	8.1	325	6	US-10-793-626-1076	Sequence 1076, App
36	70.5	8.1	689	7	US-11-113-424-46	Sequence 46, Appl
37	70	8.0	763	6	US-10-873-528-63	Sequence 63, Appl
38	70	8.0	1992	7	US-11-069-834-58	Sequence 58, Appl
39	70	8.0	2000	7	US-11-069-834-56	Sequence 56, Appl
40	69.5	7.9	178	6	US-10-467-657-2220	Sequence 2220, Ap
41	69.5	7.9	404	6	US-10-793-626-1130	Sequence 1130, App
42	69.5	7.9	574	7	US-11-024-959-300	Sequence 300, App
43	69	7.9	258	6	US-10-793-626-440	Sequence 440, App
44	69	7.9	277	7	US-11-156-084-180	Sequence 180, App
45	69	7.9	311	7	US-11-156-084-330	Sequence 330, App

ALIGNMENTS

RESULT 1
US-11-097-728-4
; Sequence 4, Application US/11097728
; Publication No. US20050260132A1
; GENERAL INFORMATION:
; APPLICANT: Erik Foehr
; APPLICANT: Sabine Muller
; APPLICANT: Daniel J. Chin
; APPLICANT: Mirella Gonzalez-Zulueta
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES DIRECTED TO RECEPTOR PROTEIN
; FILE REFERENCE: TYROSINE PHOSPHATASE ZETA
; CURRENT APPLICATION NUMBER: US/11/097,728
; CURRENT FILING DATE: 2005-03-31
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-097-728-4

Query Match	11.1%;	Score 97;	DB 7;	Length 374;
Best Local Similarity	21.8%;	Pred. No. 0.12;		
Matches	48;	Conservative 36;	Mismatches 50;	Indels 86; Gaps 11;
QY	26	YIEDNDIF----	KKILQENKLYRSSISFRYFFFDENTKKKLEYKKII-----	68
Db	156	YCFDADRFSFEAEVKGKGLRLILF-----	EVGTENLDFKALIDGVESVSRFGKQA	210
QY	69	-----TIFNLLDKGSDAIKFPVFNGLFS----	EDKVYLNNEGLLSIEI-----	114
Db	211	ALDPFILLNLLPNSTD--KYIYNGSLTSPCTDTVDWIVFKDTVSISESLAVFCEVLT	268	
QY	115	-----VKMLFFBEKNKDKFVKYSR-----	-----LDPKSFG	143
Db	269	MQQSGVYMLMDYLQNNFREQQY-KFSRQVFSYTGKEIHEAVCSSEPNVQADPENYTS	327	
QY	144	L-----YETILE-----	YDLRIADTTVHRIEDG	167
Db	328	LLVTWRPRVVDYTMIEKFAVLYQLDGEDQTKHEFLTGD	367	

RESULT 2
US-11-097-728-2
; Sequence 2, Application US/11097728

```

; Publication No. US20050260132A1
; GENERAL INFORMATION:
; APPLICANT: Erik Foehr
; APPLICANT: Sabine Muller
; APPLICANT: Daniel J. Chin
; APPLICANT: Mirella Gonzalez-Zulueta
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES DIRECTED TO RECEPTOR PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASE ZETA
; FILE REFERENCE: AGYT-006CIP
; CURRENT APPLICATION NUMBER: US/11/097,728
; CURRENT FILING DATE: 2005-03-31
; PRIOR APPLICATION NUMBER: 10/652,981
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-097-728-2

Query Match      11.1%; Score 97; DB 7; Length 2314;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

QY 26 YIEDNDIF----KKIQENKLYRSSISFRFFYDENTKKKLYKKII-----68
DB 156 YCFDADRFSSFEAVKGGKRLALSILF-----EVGTEENLDFKALIDGVESVSRFGKQA 210
QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
DB 211 ALDPPFILLNLPNST--KYIYNGSLTSPCTDTVDWIVFKDVTVSISQSLAVFCEVLT 268
QY 115 -----VKMLFFEEKNIKDEKFKVKYR-----LDPKSFGE 143
DB 269 MQQSGYVWMLDYLNQNFREQY-KFSRQVFSYTGKEEIEHVAVCSEPPENQADPENYTS 327
QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
DB 328 LLVTWERPRVVYDTMIEKFAVLVYQQLDGEDQTKHEFLTGD 367

RESULT 3
US-11-097-728-6
; Sequence 6, Application US/11097728
; Publication No. US20050260132A1
; GENERAL INFORMATION:
; APPLICANT: Erik Foehr
; APPLICANT: Sabine Muller
; APPLICANT: Daniel J. Chin
; APPLICANT: Mirella Gonzalez-Zulueta
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES DIRECTED TO RECEPTOR PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASE ZETA
; FILE REFERENCE: AGYT-006CIP
; CURRENT APPLICATION NUMBER: US/11/097,728
; CURRENT FILING DATE: 2005-03-31
; PRIOR APPLICATION NUMBER: 10/652,981
; PRIOR FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARSPLIC
; LOCATION: (1)...(2353)
; OTHER INFORMATION: PTP-zeta SM2 23a exon variant
US-11-097-728-6

Query Match      11.1%; Score 97; DB 7; Length 2353;
Best Local Similarity 21.8%; Pred. No. 1.2;
Matches 48; Conservative 36; Mismatches 50; Indels 86; Gaps 11;

```

```

QY 26 YIEDNDIF----KKIQENKLYRSSISFRFFYDENTKKKLYKKII-----68
DB 156 YCFDADRFSSFEAVKGGKRLALSILF-----EVGTEENLDFKALIDGVESVSRFGKQA 210
QY 69 -----TIFNLLDKGSDAIFPVFNGGLFS---EDKVYLNNEGLLSISEIE-----EIL- 114
DB 211 ALDPPFILLNLPNST--KYIYNGSLTSPCTDTVDWIVFKDVTVSISQSLAVFCEVLT 268
QY 115 -----VKMLFFEEKNIKDEKFKVKYR-----LDPKSFGE 143
DB 269 MQQSGYVWMLDYLNQNFREQY-KFSRQVFSYTGKEEIEHVAVCSEPPENQADPENYTS 327
QY 144 L-----YETLLE-----YDLRIADTTVHRIEDG 167
DB 328 LLVTWERPRVVYDTMIEKFAVLVYQQLDGEDQTKHEFLTGD 367

RESULT 4
US-10-793-626-3256
; Sequence 3256, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3256
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3256

Query Match      10.9%; Score 95.5; DB 6; Length 331;
Best Local Similarity 26.0%; Pred. No. 0.15;
Matches 40; Conservative 24; Mismatches 49; Indels 41; Gaps 8;

QY 26 YIEDNDIFKK----LIQENKLYRSSISFRFFYDENTKKKLYKKIIITFNLLDKGSDAI 81
DB 145 YIMTSDINHKTIEYFKQHKFYNDANHHFFKQDNIVALSEEGKLV-----LNRDGHIM 199
QY 82 KPEPVNGGLF-----SEDKVKY--LNNEGLLSISEIEEILVKML---FFEEKNI 125
DB 200 ETENGNGGVFKSLKAGYLDKMQQDHVKYIFLNN-----IDNVLVKVLDPLFAGFTV 251
QY 126 KDEKFKYKSRDPK---SFGELY-----ETLLEY 151
DB 252 TQSKDITSKTIQPKDSESVGRLVNVDCDKTVLEY 285

RESULT 5
US-11-069-642-18
; Sequence 18, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: PRAY, TODD R.
; APPLICANT: KINSELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; TITLE OF INVENTION: INHIBITING PROTEIN-PROTEIN INTERACTION
; FILE REFERENCE: RIGL-022CIF3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758

```

Db 302 GAPTGVGFSGRTKGGGPFIDSIGHRNYSKAYVSSNNDILYAEAFGIEHRVPTGV 358

RESULT 7

US-10-793-626-3298

Sequence 3298, Application US/10793626

Publication No. US20050255478A1

GENERAL INFORMATION:

APPLICANT: KIMMERLY, WILLIAM JOHN

TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

FILE REFERENCE: PU3480US

CURRENT APPLICATION NUMBER: US/10/793,626

CURRENT FILING DATE: 2004-03-04

PRIOR APPLICATION NUMBER: 60/164,258

PRIOR FILING DATE: 1999-11-09

NUMBER OF SEQ ID NOS: 4472

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3298

LENGTH: 249

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: amino acid sequence

US-10-793-626-3298

Query Match 10.2%; Score 89.5; DB 6; Length 249;

Best Local Similarity 27.3%; Pred. No. 0.34;

Matches 38; Conservative 20; Mismatches 42; Indels 39; Gaps 8;

QY 50 RYFFYDENTKKLEYKIIITFN-----LLDKGSDAIKFPVNGGLF-----SE 93

Db 18 KYFNVDANHIFHPKQDNIVALSEEGKLVLRDGHIMETPNNGSGVFKSLKKAGYLDKRWQ 77

QY 94 DKVY--LNNEGLLSISETEILVKML--FFEEKNIKDEKFVYSRLDPK---SFGELY- 145

Db 78 DHVYIFLNN-----IDNVLVKLDPLFAGFTVTSQSDKITTSKTIQPKDSSEVGLNV 129

QY 146 ----ETLEY---DLRIAD 157

Db 130 VDCKDTVLEYSELDTDIAN 148

RESULT 8

US-11-051-453-41

Sequence 41, Application US/11051453

Publication No. US20050287150A1

GENERAL INFORMATION:

APPLICANT: AMEROSINO, DONNA

APPLICANT: BABCOCK, GREGORY J.

APPLICANT: BROERING, THERESA

APPLICANT: GRAZIANO, ROBERT

APPLICANT: HERNANDEZ, HECTOR JAVIER

APPLICANT: LOWY, ISRAEL

APPLICANT: MANDELL, ROBERT

APPLICANT: MOLRINE, DEBORAH

APPLICANT: THOMAS, JR., WILLIAM D.

APPLICANT: ZHANG, HUI-PEN

TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND

TITLE OF INVENTION: USES THEREOF

FILE REFERENCE: MJ1-001

CURRENT APPLICATION NUMBER: US/11/051,453

CURRENT FILING DATE: 2005-02-04

PRIOR APPLICATION NUMBER: 60/542,357

PRIOR FILING DATE: 2004-02-06

PRIOR APPLICATION NUMBER: 60/613,854

PRIOR FILING DATE: 2004-09-28

NUMBER OF SEQ ID NOS: 82

SOFTWARE: PatentIn Ver. 3.3

SEQ ID NO 41

LENGTH: 2710

TYPE: PRT

ORGANISM: Clostridium difficile

US-11-051-453-41

Query Match 9.4%; Score 82.5; DB 7; Length 2710;
 Best Local Similarity 23.4%; Pred. No. 27;
 Matches 37; Conservative 20; Mismatches 34; Indels 67; Gaps 8;
 QY 6 ILASILEES--LIFILRI-----PFIAYIEDNDFPKILOEN 40
 DB 1747 ILVRYLEENKKILQIRIKILSNTQSFNWSIDFKDKLGLVIMSN--FKSPNSN 1804
 QY 41 KLYRSSISPRY-----PFYDENTKKLEYKKIITI-----FNL----- 73
 DB 1805 ELDRDHLGFKIDNKTYYYDESKL---VKGLININNSLFYDPPIEFNLVTGWTINGKK 1861
 QY 74 ----LKGSDAIFPVFNGGLFSEDKVKYLNNEGLLSI 107
 DB 1862 YYPDIINTGAALTSYKIINGKH-----YFNNDGVNQL 1893

RESULT 9

US-10-467-657-3876
 ; Sequence 3876, Application US/10467657
 ; Publication No. US20050260581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHIRON Spa
 ; APPLICANT: FONTANA Maria Rita
 ; APPLICANT: PIZZA Mariagrazia
 ; APPLICANT: MASIGNANI Vega
 ; APPLICANT: MONACI Elisabetta
 ; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
 ; FILE REFERENCE:
 ; CURRENT APPLICATION NUMBER: US/10/467,657
 ; CURRENT FILING DATE: 2003-08-11
 ; PRIOR APPLICATION NUMBER: GS-0103424.8
 ; PRIOR FILING DATE: 2001-02-12
 ; NUMBER OF SEQ ID NOS: 9218
 ; SOFTWARE: Seqwin99, version 1.04
 ; SEQ ID NO 3876
 ; LENGTH: 833
 ; TYPE: PRT
 ; ORGANISM: Neisseria gonorrhoeae
 US-10-467-657-3876

Query Match 9.4%; Score 82; DB 6; Length 833;
 Best Local Similarity 20.6%; Pred. No. 6.9;
 Matches 44; Conservative 41; Mismatches 65; Indels 64; Gaps 11;
 QY 1 QITQHILASILE-----ESIIFILRIFFIAYIEDN-----DIFPKILOE 39
 DB 439 KFTBELKQKILQKYPDIKPLVNLQFSDGIIDNENFAEDGYDKLKAAYPEAPFKGLDK 498
 QY 40 NKLY-----RSSISPRYFFYDE-----NTKKKLEYK-----KIITIFNLLDKGSDA 80
 DB 499 GKVSNAKDEGKDTIIMREGKYEELKALWELHKKAVLQYKIKDEAEFVDFLTAYLR-ENA 557
 QY 81 IKFPVNGGLFSEDKVKYLNNEGLLSISEIEILVKMLFPEEKNIKDEKFKVKSRLDPKS 140
 DB 558 AKFP--QAGICTAVNEAYINN-GLMLSRID-----SIEDEDFIRENTMTYRE 602
 QY 141 FGLVETLLELDRIADTTVHRIIEDGYVLRTE 174
 DB 603 FLE-----KLAQTAKIQMQLHQ-----AFYVRDE 628

RESULT 10

US-10-793-626-128
 ; Sequence 128, Application US/10793626
 ; Publication No. US20050255478A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KIMMERLY, WILLIAM JOHN
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
 ; FILE REFERENCE: PU3480US
 ; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04
 ; PRIOR APPLICATION NUMBER: 60/164,258
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 4472
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 128
 ; LENGTH: 222
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-10-793-626-128

Query Match 9.1%; Score 80; DB 6; Length 222;
 Best Local Similarity 27.0%; Pred. No. 2;
 Matches 41; Conservative 19; Mismatches 52; Indels 40; Gaps 8;
 QY 27 IEDN-DIFPKILOENKLYRSSISPRYFFYDENTKK-----KLEYKKIITIFNLLDKGSDAI 81
 DB 92 VENNIQTETKRFLYERPVR-----YDSVTPELELLKRNVDKNTVFGNLPK--KAI 140
 QY 82 KFPVNGGLFSEDKVKYLNNEGLLSISEIEILVKMLFPEEKNIKDEKFKVKSRLDPKSF 141
 DB 141 QVPKYTGTTTDPFVMI-----ETDEQDAKYLIVETK-----AENMLGDKSI 184
 QY 142 GEL-----YETL-----LEYDLRIADTTVHRIIE 165
 DB 185 GRIQKKFNTLDNLNIKYLQATSAQDVVNEIK 216

RESULT 11

US-10-793-626-106
 ; Sequence 106, Application US/10793626
 ; Publication No. US20050255478A1
 ; GENERAL INFORMATION:
 ; APPLICANT: KIMMERLY, WILLIAM JOHN
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
 ; FILE REFERENCE: PU3480US
 ; CURRENT APPLICATION NUMBER: US/10/793,626
 ; CURRENT FILING DATE: 2004-03-04
 ; PRIOR APPLICATION NUMBER: 60/164,258
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 4472
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 106
 ; LENGTH: 347
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-10-793-626-106

Query Match 9.1%; Score 80; DB 6; Length 347;
 Best Local Similarity 21.7%; Pred. No. 3.5;
 Matches 35; Conservative 32; Mismatches 56; Indels 38; Gaps 6;
 QY 31 DIFPKILOENKLYRSSISPRYFFYDENTKKLE---YKKIITIFNLLDKGSDAIFPVFN 87
 DB 125 NVAPQIERQEKMYQKMLHRSIQAOENKRIISRELHDSVIQDMLNLDVLELRLKYK--- 181
 QY 88 GGLFSEDKVKYLNNEGLLSISEIEILVKMLFPEEKNIKDEKFKVKSRLDP----- 138
 DB 182 ----HRDKV-----LAETSQRIEGLLSQLIDIRNMSVE--LRPSSLODLGIEAAFKS 228
 QY 139 --KSPGELYETLLEYDLRI-----ADTTVHRIIEDGVY 169
 DB 229 YPKQFENYGMHIKYSNIGKMRFDNEIETVVYRVVQEGVF 269
 RESULT 12
 US-10-793-626-2594

; Sequence 2594, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2594

; LENGTH: 989

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence

US-10-793-626-2594

Query Match 9.1%; Score 80; DB 6; Length 989;

Best Local Similarity 27.0%; Pred. No. 13;

Matches 41; Conservative 19; Mismatches 52; Indels 40; Gaps 8;

QY 27 IEDN-DIFKKILOENKLYRSSISFRFFYDENTKK-----KLEYKKIITIPNLLDKGSDAI 81

Db 859 VENNIOETKRFYERPPVR-----YDSVTPELELLKRNVDKNVTFEGNLPK--KAI 907

QY 82 KPFVFNGLFSEDKVYLNNEGLLSISEIEILVKMLFPEEKNIKDEKFKVYKSRLOPKSF 141

Db 908 QVPKYTGTTDPFVYMI-----ETDEQAKYLIVETK-----AENMRLLGDKSI 951

QY 142 GEL-----YETL-----LEYDLRIADTTVHRIIE 165

Db 952 GEIQKQFFNTLDNLNIKYLQATSADQVYNEIK 983

RESULT 13

US-10-793-626-3294

; Sequence 3294, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3294

; LENGTH: 371

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence

US-10-793-626-3294

Query Match 9.1%; Score 79.5; DB 6; Length 371;

Best Local Similarity 26.3%; Pred. No. 4.2;

Matches 36; Conservative 25; Mismatches 39; Indels 37; Gaps 8;

QY 24 IAVIEDNDIPK---ILOENKLYRSSISFRFFYDENTKKLYKKIIT--IPNLLDKGS 78

Db 45 IIVITNDGEWKKDNITQEIK-----NT-DELVINDVETGEISQLLSKGS 88

QY 79 -----DAIKFPVFNGLFSEDKVYLNNEGLLSISEIEILVKMLFPEEKNIK 124

Db 89 LGKSYDAV-FPLLHGNGBEDGTIOGLFVLDPYGVNGVLAASSMDKLVKQOL-FEHRG 146

QY 125 IKDEKFKVYKSRLOPKSF 141

Db 147 LPQLPYISFLRSEYEKY 163

RESULT 14

US-10-793-626-1722

; Sequence 1722, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1722

; LENGTH: 199

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence

US-10-793-626-1722

Query Match 8.8%; Score 77; DB 6; Length 199;

Best Local Similarity 26.0%; Pred. No. 3.2;

Matches 34; Conservative 20; Mismatches 57; Indels 20; Gaps 5;

QY 24 IAVIEDNDIPKILQENKLYRSSISFRFFYDENTKKLEYKKIITIPNLLDKGS----- 78

Db 56 IIVADDNKKWLQILPKKRYSMVTF-----DNKGNPLEYYFDINIKITQKNARTID 109

QY 79 ---DAIKFPVFNGLFSEDKVYLNNEGLLSISEIEE--ILVKMLFFE-EKN---IKDEK 129

Db 110 LCLDLVLVFNGEYELVDEDDLMYALQNKQISKQKQYHEAVIIAHQLMIEIEDNFSIQDKV 169

QY 130 FVKYSRLDPKS 140

Db 170 MRCYHKINHKA 180

RESULT 15

US-11-156-084-331

; Sequence 331, Application US/11156084

; Publication No. US20060010515A1

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology LLC

; TITLE OF INVENTION: Controlled expression of cytokinin biosynthetic genes leads to

; FILE REFERENCE: (38-21)

; CURRENT APPLICATION NUMBER: US/11/156,084

; CURRENT FILING DATE: 2005-06-17

; NUMBER OF SEQ ID NOS: 364

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 331

; LENGTH: 312

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-11-156-084-331

Query Match 8.8%; Score 77; DB 7; Length 312;

Best Local Similarity 24.3%; Pred. No. 5.6;

Matches 28; Conservative 23; Mismatches 46; Indels 18; Gaps 4;

QY 65 KKIIITIPNLLDKGSDAIKFPV---FNGGLFSEDKVYLNNEGLLSISEIEILV---XML 118

Db 12 KLIALLGFGSGKKSALSIELAQELDAEIFSLD-----SLSIYEDINIAKSPS 60

QY 119 FFEKNIKDEKFKVYKSRLOPKSFGELEYETLLEYDLRIADTTVHRIIEDGYYLIRT 173

Db 61 LKERRNIK-HYALDYLNIDEKNNASLFKILLEDAMRVSSKEILLIVGSSFYLKS 114

Search completed: January 24, 2006, 20:55:37
Job time : 3.78483 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 285.376 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-16
Perfect score: 6538
Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIKYDNYLLNKKIPYHK 1277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6534	99.9	1277	8	Adu98757 Borrelia
2	872	13.3	174	8	Adu98755 Borrelia
3	445	6.8	1203	8	Adv87926 Streptoco
4	445	6.8	1203	8	Adv81375 Streptoco
5	445	6.8	1203	8	Adv79179 Streptoco
6	435	6.7	1532	4	AAG90567 C. glutami
7	419.5	6.4	1279	2	Aaw98761 H. pylori
8	383	5.9	1009	8	Adu24049 Bacillus
9	382	5.8	1009	5	Aao20949 Protein o
10	382	5.8	1009	8	Adu24050 Bacillus
11	382	5.8	1009	8	Adu24023 Bacillus
12	376	5.8	1009	8	Adu24051 Bacillus
13	371.5	5.7	1644	7	Adc00979 Enterococ
14	365.5	5.6	1282	8	Adn46345 Thermococ
15	364	5.6	1629	4	Aau48711 Propionib
16	364	5.6	1629	6	Abm45230 Propionib
17	349.5	5.3	1979	3	Aab18171 Plasmodiu
18	336.5	5.1	3973	3	Aab18253 Plasmodiu
19	326.5	5.0	2295	3	Aab18180 Plasmodiu
20	316	4.8	1639	2	Aaw54145 P. falcip
21	316	4.8	1639	5	Aae29345 Plasmodiu
22	315	4.8	2184	4	Aae00425 P. falcip
23	312	4.8	1516	3	Aab18195 Plasmodiu
24	309	4.7	1654	1	Aap50777 Sequence

ALIGNMENTS

RESULT 1

ADU98757

ID ADU98757 standard; protein; 1277 AA.

AC ADU98757;

DT 24-FEB-2005 (first entry)

DE Borrelia burgdorferi antigenic polypeptide seqid 16.

KW antibacterial; vaccine; immune stimulation; immunity; antigen;

XX DNA library.

OS Borrelia burgdorferi.

PN WO2004103269-A2.

PD 02-DEC-2004.

PF 17-OCT-2003; 2003WO-US033056.

PR 18-OCT-2002; 2002US-0419401P.

PA (MACR-) MACROGENICS INC.

(TEXA) UNIV TEXAS SYSTEM.

PI Sykes KP, Hale KS, Johnston SA;

XX WPI; 2004-834155/82.

DR N-PSDB; ADU98756.

PT Immunizing a subject against Borrelia burgdorferi infection comprises

providing to the subject at least one Borrelia antigen or its fragment.

PS Claim 27; SEQ ID NO 16; 121pp; English.

The invention describes a method of immunizing a subject comprising providing to the subject at least one Borrelia antigen or its fragment to induce an immune response. Also described are: an isolated polynucleotide comprising a sequence having at least 17 contiguous nucleotides in common with a sequence not given in the specification; an isolated polypeptide having at least 5 consecutive amino acids of the sequence not given in the specification; a vaccine composition comprising at least one Borrelia antigen or at least one polynucleotide encoding a Borrelia antigen; screening for at least one test polypeptide or test polynucleotide encoding a polypeptide for an ability to produce an immune response; preparing a vaccine; vaccinating a subject; treating a subject infected

Aab18172 Plasmodiu
Abu23940 Protein e
Adp25444 Plasmodiu
Aab18265 Plasmodiu
Abr53560 Protein s
Adk64380 Disease t
Ads43855 Bacterial
Abu24510 Protein e
Abu24813 Protein e
Aam40883 Human pol
Aab18301 Plasmodiu
Aab18167 Plasmodiu
Abu19294 Protein e
Adz10441 P. gingiv
Aab18236 Plasmodiu
Aam39097 Human pol
Adq17932 Human sof
Adx06867 Cyclin-de
Ads91313 A. calcoa
Abu25330 Protein e
Aab18218 Plasmodiu

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
CC ; and detecting antibodies to a herpesvirus. The method is useful in
CC immunizing a subject against Borrelia burgdorferi infection. This is the
CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
CC useful in immunizing a subject against infection by a member of the
CC Borrelia species.
XX
SQ

Sequence 1277 AA;

Query Match 99.9%; Score 6534; DB 8; Length 1277;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKNNDIVKTNPNISLYKQSKDFIKENINKLKDFILIKNKLFSIDDNSTANTIESLL 60
DB 1 MKNNDIVKTNPNISLYKQSKDFIKENINKLKDFILIKNKLFSIDDNSTANTIESLL 60
QY 61 KYIFEELNYSVEQKAGQIEGVSVDILLFENDKDKASFNKKLKAQKQNEPIPIEDIL 120
DB 61 KYIFEELNYSVEQKAGQIEGVSVDILLFENDKDKASFNKKLKAQKQNEPIPIEDIL 120
QY 121 IIAEVKRPFTSFDAKOKLKESEDLRYLYNQYQKHGILSNKWLRYDKSKVLYGEKRY 180
DB 121 IIAEVKRPFTSFDAKOKLKESEDLRYLYNQYQKHGILSNKWLRYDKSKVLYGEKRY 180
QY 181 IEFNFKKIEBEKEKEQEWFLVFIYLRKERYLKTNSVIEVEKEQIAKEKEIIQKTLKEI 240
DB 181 IEFNFKKIEBEKEKEQEWFLVFIYLRKERYLKTNSVIEVEKEQIAKEKEIIQKTLKEI 240
QY 241 LYERPDDSIYVFKAKNIYDEKFKLSGKEITQHILASLIESIIFILRIPIAYIENDIF 300
DB 241 LYERPDDSIYVFKAKNIYDEKFKLSGKEITQHILASLIESIIFILRIPIAYIENDIF 300
QY 301 KKILQENKLYRSSISFRPYFYDENTKKLEYKIIITIFNLLDKGSDAIKPPVFNGLFSE 360
DB 301 KKILQENKLYRSSISFRPYFYDENTKKLEYKIIITIFNLLDKGSDAIKPPVFNGLFSE 360
QY 361 DKVKYLNNEGLLSISEEILVKMLPPEEKNIKDEKPKVSRDLPKSGFGLYETLLEYDL 420
DB 361 DKVKYLNNEGLLSISEEILVKMLPPEEKNIKDEKPKVSRDLPKSGFGLYETLLEYDL 420
QY 421 RIADTTVHRIIEBGVYLIIRTEEELENKKVANKIATYLGNIYLTSRSDLRKSGAYTTPDD 480
DB 421 RIADTTVHRIIEBGVYLIIRTEEELENKKVANKIATYLGNIYLTSRSDLRKSGAYTTPDD 480
QY 481 LTDFPMVISSIEEQLKTKSPIDIKIINDNSCGSGHPLISCLDYLTCKWYELDKFEDVKKEL 540
DB 481 LTDFPMVISSIEEQLKTKSPIDIKIINDNSCGSGHPLISCLDYLTCKWYELDKFEDVKKEL 540
QY 541 DEEYRVIIERSEEDVDQDSISKELVLRMLLKXCIYGVNDINPISVEITMLSLWINTPIFG 600
DB 541 DEEYRVIIERSEEDVDQDSISKELVLRMLLKXCIYGVNDINPISVEITMLSLWINTPIFG 600
QY 601 TPLSFIEHHIKAGNALGYTKDEFDIVKKKFSFGPSLFFKKRIKEIITILEDIYQKIKGI 660
DB 601 TPLSFIEHHIKAGNALGYTKDEFDIVKKKFSFGPSLFFKKRIKEIITILEDIYQKIKGI 660
QY 661 NDTTKEDIEKSKYIKEYEESKDIDLRIIFSLIKLYSLSPDKSLNMEFSDIASVISLIE 720
DB 661 NDTTKEDIEKSKYIKEYEESKDIDLRIIFSLIKLYSLSPDKSLNMEFSDIASVISLIE 720
QY 721 NILGNKTSSEDEKIEKIRKLSSYYKGFHYGIEPDPDIOEGPDIVIGNPPWEKTKFNETEF 780
DB 721 NILGNKTSSEDEKIEKIRKLSSYYKGFHYGIEPDPDIOEGPDIVIGNPPWEKTKFNETEF 780
QY 781 FSKHIPNRYKLGIEQNIIEKEILSKONHPLSIEWNEEKNSIIAINNIYKDFKCFKTSGG 840
DB 781 FSKHIPNRYKLGIEQNIIEKEILSKONHPLSIEWNEEKNSIIAINNIYKDFKCFKTSGG 840
QY 841 DPNLFRYFVTNFKLIEKGNLTYLVPISAINNESSRILRHIFARYKLANIYGFENKRR 900
DB 841 DPNLFRYFVTNFKLIEKGNLTYLVPISAINNESSRILRHIFARYKLANIYGFENKRR 900

QY 901 FKDVHSSFKFAIFQLSNIKESTSSPKAFPMIQSSNLIKETRLDKSDDAYKGIELNI 960
DB 901 FKDVHSSFKFAIFQLSNIKESTSSPKAFPMIQSSNLIKETRLDKSDDAYKGIELNI 960
QY 961 NQIKKLSPIQESIIIEFKONEEFTLINKMPKSPFSGALGEGYIDFKKGLDPSIKNRKSLKKEC 1020
DB 961 NQIKKLSPIQESIIIEFKONEEFTLINKMPKSPFSGALGEGYIDFKKGLDPSIKNRKSLKKEC 1020
QY 1021 NKNKLIFFLYSGANIHFQNSRPFEDKAKESSKLLMTDKEDLEKVLTKDNOYOTERVFYRA 1080
DB 1021 NKNKLIFFLYSGANIHFQNSRPFEDKAKESSKLLMTDKEDLEKVLTKDNOYOTERVFYRA 1080
QY 1081 IASNTNERTMISTLSPGNCYCNSIYINDEKTPISLYKKLFIISFNSFVDFLLRRFVD 1140
DB 1081 IASNTNERTMISTLSPGNCYCNSIYINDEKTPISLYKKLFIISFNSFVDFLLRRFVD 1140
QY 1141 SNVLKSCLYQCPMPQPEEKEILSNPLYLNLAQNTSLIAKNDPENPKYLLYLEYFKFDKE 1200
DB 1141 SNVLKSCLYQCPMPQPEEKEILSNPLYLNLAQNTSLIAKNDPENPKYLLYLEYFKFDKE 1200
QY 1201 KVNKILKLDKEDEFFKEKENENNFIASLYSLAKEDFITLLGDFKALKNKKKGEDYISSL 1260
DB 1201 KVNKILKLDKEDEFFKEKENENNFIASLYSLAKEDFITLLGDFKALKNKKKGEDYISSL 1260
QY 1261 IKGYDNYLLNNKIFVHK 1277
DB 1261 IKGYDNYLLNNKIFVHK 1277

RESULT 2

ADU98755
ID ADU98755 standard; protein; 174 AA.
XX
AC ADU98755;
XX AC
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 14.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
OS Borrelia burgdorferi.
XX
XX WO2004103269-A2.
PN
XX 02-DEC-2004.
XX
XX 17-OCT-2003; 2003WO-US033056.
XX
XX 18-OCT-2002; 2002US-0419401P.
PR
XX (MACR-) MACROGENICS INC.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KP, Hale KS, Johnston SA;
XX
XX WPI; 2004-834155/82.
DR N-PSDB; ADU98754.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises
XX providing to the subject at least one Borrelia antigen or its fragment.
PS Claim 27; SEQ ID NO 14; 121pp; English.
XX

CC The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;

CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.

XX
 SQ Sequence 174 AA;

Query Match 13.3%; Score 872; DB 8; Length 174;
 Best Local Similarity 99.4%; Pred. No. 6.7e-43;
 Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 268 EITQHILASILESIIFILRIFFIAYIEDNDIFKKILQENKLYRSSISFRFFYDENTKK 327
 DB 1 QITQHILASILESIIFILRIFFIAYIEDNDIFKKILQENKLYRSSISFRFFYDENTKK 60
 QY 328 KLEYKKIITIFNLLDGGSDAIFPVPNGGLFSDKVKYLNNEGGLLSISIEIEILVQMLPF 387
 DB 61 KLEYKKIITIFNLLDGGSDAIFPVPNGGLFSDKVKYLNNEGGLLSISIEIEILVQMLPF 120
 QY 388 EKNINDEKFKVYSRLDPKSPGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 441
 DB 121 EKNINDEKFKVYSRLDPKSPGELYETLLLEYDLRIADTTVHRIEDGVYLIRTE 174

RESULT 3

ADV87926
 ID ADV87926 standard; protein; 1203 AA.

AC
 AC ADV87926;

XX
 XX 24-FEB-2005 (first entry)

XX Streptococcus agalactiae protein sequence, SEQ ID 320.

XX Antibacterial; Vaccine; bacterial infection.

XX Streptococcus agalactiae.

XX FR2824074-A1.

XX 31-OCT-2002.

XX 26-APR-2001; 2001FR-00005642.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;

XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst P;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips
 PT and identification of therapeutic targets.

PS Claim 6; SEQ ID NO 320; 2687pp; French.

XX The present invention relates to novel Streptococcus agalactiae
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.
 CC agalactiae involved in the synthesis of amino acids, cell membranes,
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and
 CC phospholipid metabolism, nucleotide metabolism including purines,
 CC pyrimidines and/or nucleosides, regulatory functions, replication,
 CC transcription, translation, protein transport, adaptation to atypical

CC conditions, sensitivity to medicines and/or analogues, functions related
 CC to transposons, biosynthesis of cofactors, prosthetic groups and
 CC transporters, cell membrane proteins and cellular machinery. (I) are
 CC useful for the detection and/or amplification of nucleic acids.
 CC Pharmaceutical composition comprising (I) or (II) are useful for
 CC treatment of a bacterial S. agalactiae infection. Note: W0200292818A2 is
 CC equivalent for the present basic patent FR2824074A1. W0200292818A2
 CC contains 6617 sequence whereas the present patent only contains 2344
 CC sequences.

XX
 SQ Sequence 1203 AA;

Query Match 6.8%; Score 445; DB 8; Length 1203;
 Best Local Similarity 19.4%; Pred. No. 4.3e-17;
 Matches 285; Conservative 241; Mismatches 400; Indels 546; Gaps 71;

QY 26 KKENIN-KLKDPFILLIKNNKLSIDNSTEAN-----IESLLKYIFEEINLTSVEQOKAG 77
 DB 25 KEKVFQKLVSVLTHLKD-----ESESSEYQKLLKVFLESVLPYNF--INTS----- 71
 QY 78 QIEGVESRYDILLFENDKDKASFNKMLKEAK---KNEPIPIEDILIIAEVKRPTFSFDA 134
 DB 72 -----SRIDLAIY-NGKDANSLSGLIPCKSLFNKSEMMSTEKI----- 109
 QY 135 KDKLSESEDLRYLNQ--YQHYG-----ILNGKVRWLYDK---SKVLYGEKRYL----- 181
 DB 110 --NSKAQFIVVYVYIQLERLFNKNLEIKKCIITNGLSWFVIEAKEPEKHFKNKGLVDLVT 167
 QY 182 -----EFNFKIEEKEEYKEQEFVLFYILIRKERYLKTNSVIEVEKEQIAKE 229
 DB 168 KFRNQLSNSKTDLYSEVIAPEIDKAFKGIIVIAHFDSLQALVKTSKSIIEIKNNLTQL 227
 QY 230 KEIIQKTLKEILYERPDSDIVF---KIAKNYDKEFKLSGKEITOHILASILE----- 279
 DB 228 YRFF--TAENLANKE-----IFTDSNKNKNFYDELLYLMGLEETSGYTKIISRLKPIK 280
 QY 280 -----ESIIFILRIFFIAYIEDNDI-----FKILOEN-KLYRSSISFRY 318
 DB 281 RQYSFVENIINKLEMKOVSKQEDIAIQLTAVVNTNRILFLKLLLESQVLFNKDESRYF 340
 QY 319 FFYDENTKKLEVKKIITIF-----NLLDKGSDAI--KF---PVFNGGLFSEDKVKYLVAN 368
 DB 341 LTYE---KLPNFEIYGLFFPAVLAKKVSERNDRVQKEGYVPYLNSSLFEETEIE-ISR 395
 QY 369 EGL-----LSISBIEEILVKMLFFEBKNIK--DEK-----FVKYSRLDPKSPGELYETL 415
 DB 396 DGIIDRLPEGDIE-----IFSXTALAGVDKRRKGNINIEY-----LFEFL 438
 QY 416 LEYDLRIADTTVHRIEDGVYLIRTEELENNKKNKIATYLGKNIYLTSRSLDRKSGAY 475
 DB 439 DSYDFSTS-----ISHHEKSKNDLIN--ASVL-GLIF---EKINGYRDSGF 478
 QY 476 YTPDDLDTDFM-----VISSIEEQKTKSPL-DIKI 504
 DB 479 YTPGNITMYMSKKAIRTAADVKNELLGNCTVETBEIKFAIGHSVENARKVSQAIDDLKV 538
 QY 505 IDNSCGSGHFLISCL-----DYLTEKVMYVELDKFEDVKCKELDEE--- 543
 DB 539 CDPVAGSGHFLVLSILNEIITALKSDNLIFDADGNVIGNLIQCVINDELVIQDMGNNFL 598
 QY 544 YRVIIIESEEDVQDQSIKELVKRLMKLKKCYIGVDINPISVEITMLSLMI-----NTFIF 599
 DB 599 YQTGNQOQSEQ--IQKAIFNQ---KRHILENSLFAVDINPSSVNICRLRLWIELLKSSVYV 653
 QY 600 G-----TPLSFIEHHIKAGNALLGYTKDEF---FDIVKKPESGSLFK----- 640
 DB 654 QDITDNTQNPVLTLPNIDINIKVGDSSL--HKFEFDYFDMRKTDKDYLSLVKDYKETNN 711
 QY 641 KRIKEIITILEDIYOKINGINDTTKEDIEKSKYKVEESKDIDNL--RIIFSLIKLYSL 699
 DB 712 KKKV-----ABIWEK-----IERLKHSDDTTASSPELKSLSLIERNKAGAI 754
 QY 700 SFDKSLNMEFSDIASVISLIENILGNKTSSEDEKIEKIRKLSYYKFFHYGIEFPDIOE 759

Db 68 LNNENNE---VEVITIEFKALKNPNEFIKGGDLNVKAFHESLLSYLTERKEGNNNLKHL-IL 123
 Qy 160 SNQVWRLDYKSV-LYGEKRYIEFNFEXKIEEKEEYKEQWFLFIY-----LIRKERY 212
 Db 124 ATIKELYIIDANEFVFNKDKIEENAFKPNCHDR---KGNDRTRYKAFYDACQKRLNEBFRS 180
 Qy 213 LKTSNVEVEKEQIAKEKEIIQTKLEILYERPDSDIVFKIAKNIYDKFEFKSGKE----268
 Db 181 LK-YHYIPLKENLAL---IYQALSFPNFKLIKPYSDANTLNKDFEELLYILGLEQND 236
 Qy 269 -----ITOHILASIL-----BESIIFIL-----RIFPIAYIEDNIFKXILQE 306
 Db 237 KGKILIKPSRTONSLSDALKKEYKNLDDDEVALLIANNRILFLRLLES-----286
 Qy 307 NKLYRSISIRYFFYDEN-----TKKLEYKKIITIP---NLLDKGSDAI-----KF 350
 Db 287 -----LLISFKHF---ENPFLTTNFENFNDLNTLFFVLAKKONSRLPEIKEDKILEKI 338
 Qy 351 PVFNGGLFSE-----DKVYLNNEGLLSISIEIEILVKMLPFEEKNIKDEK---FVKY 400
 Db 339 PYLNSLFDKTPLELKGHEIKLDNKKL-----EYKNSVLKKDKQYKQKPLPLKY 391
 Qy 401 SRLDPKSGFGLYETLLEYDIRIADTTVHRIIEBGVYLIRTEELENKKNKIATYLGNI 460
 Db 392 -----LFKFLRLYKF---TTTPDKDKONT---DTSESRLINPSV-----L 425
 Qy 461 YLRSRLDRKSGAYTPDDLTDFPWISSIE-----BQLKT-----496
 Db 426 GLVFEKLVNGYKESFYTPSPITSYMKCESITPIVLDFKFNAYQWDCENLKALGEIDRNF 485
 Qy 497 -----KSPIDIKIINDSCSGHFLISCLDVLTEKWKWYELDKFE-----DVKKELDE 542
 Db 486 SNEKAKYMLTLRLICDPAVSGHFLVSALEMWV-RVAYELGLTASLYRYDLKLENDE 544
 Qy 543 -----BYRVIIIESEYDQDSISKELV-LKRMLLKXCIYGVNDINPISVEITMLSL 592
 Db 545 IITHHTPTGEIFVNYKPDSENDPHHIQKELFNKKSIIENCLFGVDINPNSCEITKRL 604
 Qy 593 WI-----NTPFPGTPLSFIEHIIKAGNAL-----ILEDIYQKI-----LGY 619
 Db 605 WIELLYSYIYIFBKGNVNALET-LPNIDINIKANSLSIRFALKDKALKSEKNKNLEY 663
 Qy 620 TKDEFDPDVKKKPSGSLFK-KRIKEIIT-----ILEDIYQKI-----657
 Db 664 SIABYKELVK-----IYKDPKILETLTHPIKOSNAVKYAKERYIQELKQNPKNDFK 715
 Qy 658 KGINDTTKEDIEKSKTIYK---EYBESKDIDNLRIIFSLIKLYSLFDKSLN---MEFSDI 712
 Db 716 KALNDR---IEKIKAFKLTLPNPPKELFKFKLKEHLELYGKSIIEEANYNGLELEAL 771
 Qy 713 ASVISLIENILGNKTSSEDEKIEKIRKLSYKF-----IRQSHI-KDL 861
 Db 772 A-LEKOMANLFFDYRYPFKLDSKVVGLEHFNRYLTSYKDLQDENERYANALEWRFEF 830
 Qy 755 PDIOE-----GPDIVGNPPWEKTKNETEFSKHIPNYRKGLIKEQNIKOBILSKDN 808
 Db 831 PEVLDDGSGFGDCIIGNPPY-----IRQSHI-KDL 861
 Qy 809 HP-LSIEYNEEKNISIIANNIYKDPKCTSGGDPNLFYFVTFNLKLIKEKGNLTLYVP 867
 Db 862 KPLLEKYQDPYNS-----TADIYTYFFALAFHLLKKEGFSAFITS 902
 Qy 868 SAIWNESSRILRKHIFAR-----YKLNYYQENKPKFQDVHSSKPAIFQLSNIKES 921
 Db 903 NKYTRAKYGAKLREWLLKTKTIVSYMBELNALKVPESAA---VDTSTI-----IHFIKQT 952
 Qy 922 TSSPKAFMIQSSNLIKETIRDLKDSKDAYKGIELNINQIKLSPIQSIITEFKDNEE 981
 Db 953 PSK-----ESEFKYETPTNDKDLKSTPHLLMKQNVLSTESFIFANATILDLRD---1002
 Qy 982 FTLINKMFSKPSALGEGYIDPKGLDPSIKNRKSLKKECNKNLIFLYSGANIHQFNSRF 1041
 Db 1003 -----KTESVGTPLKDWIDIQINYGIK-----TGA-----NEAF 1030

Qy 1042 PEDDAKSSKLLWIDKEDLEKVLTKDNOYQOTERVFY-----RAJASNTNERTMTSLSP 1096
 Db 1031 IIPTEKRE-----EILNACKTOEERETERLIKPILRGDKDIKRYSYEWAHLWVINT 1081
 Qy 1097 GNCYCVNSIYINDEKTPISLYKKLPIISFNFSFVDFLLRRFVDSNV---LKSCLYQCPM 1153
 Db 1082 HNGYTSS---LKSIPPIDIEKYPAKHAHLDAAH-YDTIATRCDOGDTVPYHLRNCAYL---1134
 Qy 1154 POPEKE-----ILSNPLYLNLAKE-----NTSLLI-----1178
 Db 1135 -EDFEKEKIWASVGVEYCMIPGLLILDTNYFEVSKFEGNTKNYLLGLLSKLLTFWLK 1193
 Qy 1179 AKNDP-----ENFKY-LLYLEYKFK---DKEKNVKILKLDKDEFFKEK-----1218
 Db 1194 AKNTPLGDMGAVRNTKINMELPWKITAKNKKIADKIILAVDKILQAKEKDPKANTQKL 1253
 Qy 1219 ENENNFITIASLYSLAKED 1236
 Db 1254 EKEIDALVYQLYHLTDEE 1271
 RESULT 8
 ADU24049
 ID ADU24049 standard; protein; 1009 AA.
 AC ADU24049;
 XX
 DT 27-JAN-2005 (first entry)
 DE Bacillus pumilus BpmI restriction endonuclease mutant protein (D74A).
 KW Type IIG restriction endonuclease; molecular cloning;
 KW Gene characterisation; BpmI restriction endonuclease; BpmIRM; mutant;
 KW mutein; enzyme.
 XX
 OS Bacillus pumilus.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 74 /note= "Wild type Asp is substituted by Ala"
 FT
 XX US2004219584-A1.
 XX
 PD 04-NOV-2004.
 XX
 XX 15-MAR-2004; 2004US-00800946.
 XX
 XX 20-OCT-2000; 2000US-00693146.
 PR 17-MAY-2002; 2002US-00150028.
 PR
 XX (XUSS/) XU S.
 PA (KOB/) KOBBE D.
 PA (ZHUZ/) ZHU Z.
 PA (SAMU/) SAMUELSON J.
 XX
 PI Xu S, Kobbe D, Zhu Z, Samuelson J;
 XX WPI; 2004-783790/77.
 XX
 PT Altering specificity of Type IIG restriction endonuclease by transforming
 PT cell with DNA encoding catalytic domain and methylase N-terminal of first
 PT endonuclease ligated to methylase C-terminal and specificity domain of
 PT second endonuclease.
 XX
 PS Example 5; Page; 34pp; English.
 PS
 CC The present invention relates to a method of altering the cleavage
 CC specificity of a type IIG restriction endonuclease (E) having cleavage,
 CC methylase (M) and specificity domains. The method involves transforming
 CC cell with DNA encoding catalytic domain and methylase N-terminal of first
 CC endonuclease ligated to methylase C-terminal and specificity domain of

XX Xu S, Kobbe D, Zhu Z, Samuelson J;
 XX WPI; 2004-783790/77.
 DR N-PSDB; ADU24022.
 XX
 XX Altering specificity of Type IIG restriction endonuclease by transforming
 PT cell with DNA encoding catalytic domain and methylase N-terminal of first
 PT endonuclease ligated to methylase C-terminal and specificity domain of
 PT second endonuclease.
 XX
 XX Disclosure; SEQ ID NO 4; 34pp; English.
 XX
 XX The present invention relates to a method of altering the cleavage
 CC specificity of a type IIG restriction endonuclease (E) having cleavage,
 CC methylase (M) and specificity domains. The method involves transforming
 CC cell with DNA encoding catalytic domain and methylase N-terminal of first
 CC endonuclease ligated to methylase C-terminal and specificity domain of
 CC second endonuclease. The method is useful for cleaving DNA molecules into
 CC small fragments for molecular cloning and gene characterization and for
 CC overproducing restriction endonucleases by genetic manipulation. The
 CC present sequence is the *Bacillus pumilus* BpmI restriction endonuclease
 CC (BpmIRM) protein.
 XX
 XX Sequence 1009 AA;
 SQ
 Query Match 5.8%; Score 382; DB 8; Length 1009;
 Best Local Similarity 19.4%; Pred. No. 1.6e-13;
 Matches 261; Conservative 199; Mismatches 417; Indels 468; Gaps 55;
 14 ISLYKQLSKDPKKE-NINKLKDPFI--LIKNNKLFSDDDNSTEANTIESLLKLYFEELNYS 70
 7 VDKTKAHRSTFLPTNETQNDNFIDPLKSLGWDVNTKGTHTI---LRDVIQEYIE 63
 71 VEQKQACQIEGVSRVDILLFENDKOKASFNNKLEAKKNEPIEDILIIAEVKRPTF 130
 64 IKD-----EETKKNPDYTLRINGTRK-----LFVEVKKPSF 94
 131 SFPAKDKLKESED--QYRYLVNQKHGILSGKWRLYDKSLVKLYGEKRYIEFNFXXI 188
 95 NI-----LKSAAAFQTRRYGWSANLIGSVLTNPFHLVIYD-----CRY-----T 134
 189 EEKEEYKEQWFLVFIYLRKRYLTSNIVIEVEKEQIAKEKIIQKTLKEIILVERPDD 248
 135 PDKSDNEHIAKYVFSVEEYEPADETKDIISYESANSALDEMFVNTNR--VGETPDEY 192
 249 IVFKIAKNYDKGFKLSGKEITQHILASILEESIIFIL-----RIFFIAYIEDNDIFKKI 303
 193 FLOQI-ENWREKLAKTAIKNTE-----LGEEDVNFIVQRLNRIIFLRVCEORTIEKY- 245
 304 LQENKLYRSSISPRYPFYDENTKKLYKKIITIFNLLDKGSDAIKFPVFNGLFSEDKV 363
 246 -----ETIKSIKNYEBLKOLFQKSDR-----KFNGLFD----- 274
 364 KYNNEGLISTSEIEELVXK--LFFEKNKIDKEPKVYSRLDPKSGFELYETLLBYDL 420
 275 -FIDDTLLLEVDNSNVLIEFSLYFPQP-----YDFSVDPTILSQYERFLQGEI 327
 421 RIADTTVHRITEDGVILIRTEEELENKKVKNKIATYLGNIYLTSRSLDRKKSGAYTTPDD 480
 328 -----LIESG-----GTHITESPEVAASNGVVPFKI 355
 481 LTDFMV---TSSIEQLKTSPLDIKIIDNSCGSHFLISCLDYLTEKVMYLDKDFEDVK 537
 356 IVEQIVKDTLPTLPTLTKGKFNELCNLKIADICCGSGTELLISSYDPLVEKV-----ME 406
 538 KELDEEVRVIEESE-EYDVQDSISKELVLEKMLKKCIYGVNDINPISVEITMLSLWINT 596
 407 KILEEN-----IDDSLDVYETEEGILTLKAKRNILENNLFGVDVNPYAVEAFSEF----- 458
 597 FIFGTPLSFIEHHIKAGNALLGYTKDBFFDVIKKKFSFGSLFKKRIKEIITILEDIYQK 656
 459 -----LKKLE----- 464

QY 657 IKGINDTTKEDIKSKYIKEYESKDIDNLRIFSLIKLYSLSPDKSLNMBFSDIASVI 716
 DB 465 --GENEASVNN-----FIHEHE-----DKIL-----PDLSII 490
 QY 717 SLIENILGNK-----TSSDEKKEIKIRKUSYKFFHYGFPPDIQIE--GFDVIGNP 768
 DB 491 KCGNSLVNDKPPFPFESLEDEILFKANP-----FEWESEFPDIMGANGFDALIGNP 543
 QY 769 PHEKTFNETEFSSKHIPNYRKLGIKEQNIYKQETLSKDNHPLSIEYNEEKNSILAINNI 828
 DB 544 PY-----VRIQNM-----KKYSPEEIEYQSKDSE----- 568
 QY 829 KYKDFKCTGGDPLNLFYVTFEN--LKLIEKGNLTLYLPSAINWNESSRILRKHIFAR 886
 DB 569 -----YVAKETVDKFLFIEBALILNLTGLLGIIPHKFFITKGGKELRFAIEK 621
 QY 887 YKLIYIYQFENKRFKDVHSSFKFAIFQLSNIKESTSSPKAFMIQSSNLIKBITRDLK 946
 DB 622 HQISKIINFVTVQPPGRATYTAIIIOANKM-----AQFYKVKVSNISAEVL----- 669
 QY 947 DSKDD--AYKGIELMINQIKLSPIQESIEEFKDNNEEFTLINKMPSKFSALGEGYIDFKK 1004
 DB 670 DSEENTCVYSSEKYNSDPWFILSPETEA-----FTKFTF--AQPEKLGIE-ITDISV 718
 QY 1005 GL-----DPSIKNKRKSLKKE-----C-----NNKLI 1026
 DB 719 GLQTSADKIYIPIFENETSDTYIFNVKGRYIEIKSICCPAIDYDLSFGSFESIQGNAQMI 778
 QY 1027 FLYSGANIHQFNSRFFEDKAKESKLLWIDKEDLEKVLTKDNQYQTERVYFRAIASNTN 1086
 DB 779 FFY---EIRDEBAYLLEETLENDYPLAWNLYNEPKEALEKSLQGRNPKWYQGRSQSL 835
 QY 1087 ER-----TWISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISINFSVDFELLRR 1137
 DB 836 SKFHDKEKLITWLTATKPPYVLDNRNLLFTGGNGP-----YYGLINQSIYSLHYF----- 886
 QY 1138 FVDSNVLSKSLYQCPMPQPEEKEIILSNPLYNLAKNTSLIAKNDPENFKYLLYLEVFKF 1197
 DB 887 -----LGILSHPVIESNVKARA-----SEFRG-----SYSH 913
 QY 1198 DKEKVNKI---LKLDKEDFEFFK-----EK-----ENENN----- 1223
 DB 914 GKQFIKIPRKIDPDQDEVDKYNVTVTVEKLIITTDRIKSESNGPRRRMLRRRLDAL 973
 QY 1224 -----PIIASLYSLAKEDFITLLGD 1243
 DB 974 SNQLIQVINELYNISDEEYTTVLND 998
 RESULT 12
 ADU24051
 ID ADU24051 standard; protein; 1009 AA.
 XX
 AC ADU24051;
 XX
 XX 27-JAN-2005 (first entry)
 DT
 DE *Bacillus pumilus* BpmI restriction endonuclease mutant protein (E88A).
 XX
 XX Type IIG restriction endonuclease; molecular cloning;
 KW gene characterisation; BpmI restriction endonuclease; BpmIRM; mutant;
 KW mutagen; enzyme.
 XX
 OS *Bacillus pumilus*.
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 88
 FT /note= "Wild type Glu is substituted by Ala"
 XX
 PN US2004219584-A1.
 XX

PD 04-NOV-2004.
 XX 15-MAR-2004; 2004US-00800946.
 XX 20-OCT-2000; 2000US-00693146.
 PR 17-MAY-2002; 2002US-00150028.
 XX (XUSS/) XU S.
 PA (KOB) KOBBE D.
 PA (ZHUZ/) ZHU Z.
 PA (SAMU/) SAMUELSON J.
 XX Xu S, Kobbe D, Zhu Z, Samuelson J;
 PI WPI; 2004-783790/77.
 XX
 PT Altering specificity of type IIG restriction endonuclease by transforming
 PT cell with DNA encoding catalytic domain and methylase N-terminal of first
 PT endonuclease ligated to methylase C-terminal and specificity domain of
 PT second endonuclease.
 XX
 PS Example 5; Page: 34pp; English.
 XX
 CC The present invention relates to a method of altering the cleavage
 CC specificity of a Type IIG restriction endonuclease (E) having cleavage,
 CC methylase (M) and specificity domains. The method involves transforming
 CC cell with DNA encoding catalytic domain and methylase N-terminal of first
 CC endonuclease ligated to methylase C-terminal and specificity domain of
 CC second endonuclease. The method is useful for cleaving DNA molecules into
 CC small fragments for molecular cloning and gene characterization and for
 CC overproducing restriction endonucleases by genetic manipulation. The
 CC present sequence is the Bacillus pumilus Bpmi restriction endonuclease
 CC mutant protein (B88A). Note: This sequence is not shown in the
 CC specification but is derived from Bacillus pumilus Bpmi restriction
 CC endonuclease wild type protein shown as SEQ ID NO:4 in page 16-18 of the
 CC specification.
 XX
 SQ Sequence 1009 AA;
 Query Match 5.8%; Score 376; DB 8; Length 1009;
 Best Local Similarity 19.3%; Pred. No. 3.6e-13;
 Matches 260; Conservative 199; Mismatches 418; Indels 468; Gaps 55;
 QY 14 ISYKQLSKDFIKK-NINKLDPFI--LIKNLFSIDNSTANIESLKYIFELNYS 70
 DB 7 VDKYKAHRSTFLKPTNETQLRNFIDPLKSLGWDVNTKGTTHI---LRDVIQEEYIE 63
 QY 71 VEQQKAGIIEGVESVDILLFENDKDKASFNKLEKAKKNEPIPIEDILIIAEVKRPTF 130
 DB 64 IKD-----EETKKNPDYTLRINGTRK-----LFVAVKKPSF 94
 QY 131 SPDAKDKLKESED--QLYRYLNOYQKHVGLSKVWRLYDKSKVLYGEXRYIEFNFKXI 188
 DB 95 NI-----LKSAAKAFQTRRGWGANLGSIVTWFHVIYD-----CRY-----T 134
 QY 189 EEKEEYKEQWFLFYILIRKRYLKTNSVIEKEQIAKEBIQKLEILYERPDSS 248
 DB 135 PDKSDNEHARYKVFSEYEEAFDEIKDIISYESANSALDEMDVNTNR--VGETFDEY 192
 QY 249 IVFKIANIYDKRPFKLSGKGIHQHILASILEESIIFL-----RIPFIAYIENDNFKKI 303
 DB 193 FLOQI-ENWREKLAKTAKNTE-----LGEEDVNFIVQRLNRIIFLRVCEDRTEKY- 245
 QY 304 LQENKLYRSSISFRYPFYDENTKKLEYKIIITENLLDKGSDAIKFPVNGGLFSEDKV 363
 DB 246 -----ETIKSKNYEELKOLFQKSDR-----KFNGLFD----- 274
 QY 364 KYNLNEGLGISSEIEELVRM--LFFEEKNKDEKVKYSLRDPKSGFELYETLLEYDL 420
 DB 275 -FIDDTLLLEVEDSNVLIIFSDLYPQSP-----YDFSVDPTILSQIYERFLQEI 327
 QY 421 RIADTTVHRHIEGCVYLIRTEBELENKKNVKNKIATYLGNIYLSRSRDLRCKSGAYTTPDD 480

DB 328 -----IIESG-----GTFHITESPEVAASNGVPTPKI 355
 QY 481 LTDPMV---ISSIEQLTKSPLDYKIIDNSCGSHFLISCLDYLTPEKVMYELDFEDVK 537
 DB 356 IVEQIVKDTLTPLTEGKFNELCNLKIADICCGSGTFLISSYDFLVEKV-----ME 406
 QY 538 KELDEYRVIIIESE-EYDVQDSISKELVLRMLLKXCIYGVNDINPISVEITWLSLWINT 596
 DB 407 KIIEN-----IDSDLVYETEBGLILTLKAKRNILENNLFGVDVNPYAVEAEFSL---- 458
 QY 597 FIFGTPLSFIEHRIKAGNALLGYTDEFFDIVKKKPESGSLFKKRIKIITILLDIYQK 656
 DB 459 -----LLKLE----- 464
 QY 657 IKGINDTTKEDIEKKIYKYEESKDIDLRIIFSLIKLYSLSPFKSLNMFESDASVI 716
 DB 465 --GENEASVNN-----FIEHE-----DKIL-----PDLTSII 490
 QY 717 SLIENILGNK-----TSSEDEKIEKIRKLSYKFFHYGIEFFDIQZ--GFDIVGNP 768
 DB 491 KCGNSLVDNKKFFEFMPESLEDEILFKANP-----FEWEEFFDIMANGGDFAIIGNP 543
 QY 769 PWEKTFNETEFSSKHIPNYRKLGIKEONIIOEILSKDNHPLSIEYNEEKNSIIAINNI 828
 DB 544 PY-----VRIQNM-----KKYSPEEIEYYQSKDE----- 568
 QY 829 YKPDFKCFSTGGDPLNRYFVFTN--LKLIEKGNLTLYLPSAIMNNESSRILRKHIFAR 886
 DB 569 -----YTVAKETVDKVFIERALILNLTGLLGYIIPHKFFITKGKELRKFIAEK 621
 QY 887 YKLYIYQENKRRKDVHSSFKFAIFOLSNIKESTSSFKAKFMIOSSDNLKEITRDLK 946
 DB 622 HOISKIINFGVTQVPPGRATYTAILLIQANKM-----AOPKYKKNVSNISAEITL-- 669
 QY 947 DSKDD--AYKGLTELNIQIKLSPIQESIIIEPKDNEETLINMKFSGSALGSGYIDFKK 1004
 DB 670 DSEENTCVISSEKYNSDPMIFLSPETEAV-----FTKFTB--AQFEKLG-ITDISV 718
 QY 1005 GL-----DPSIKNRKSLKE-----C-----NNKLI 1026
 DB 719 GLQTSADKIYIFIPENETSDTYIFNYKGRYEIEKSCCPAIVDLSFGSPESIQGNAQMI 778
 QY 1027 FLYSGANIIHQFNRFPEDKAKESKLLWIDKEDLEKVTKONQYQTERVFFYRAIASNTN 1086
 DB 779 FPY---EIRDEEAYLLEETENDYPLAWNLYNEFEKLEKSLQGRNPKWYQYGRSOSL 835
 QY 1087 ER-----TWISTLSPGNCVCNSIYINDEKTPISLYKKLFIISFNSFVDFELLR 1137
 DB 836 SRPHDKELIWTVLATKPPYVILDRNNLLFTGGNGP---YYGLINQISVLSLHYF----- 886
 QY 1138 FVDSNVLSCLYQCPMPQPEEKEILSNPLYLMLAKNTSLIIAKNDPENFKYLLYLYBFKF 1197
 DB 887 -----LGILSHPVIESNVKARA-----SEFRG-----SYSH 913
 QY 1198 DKEKNKI---LKDKEDPEFK-----EK-----ENENN----- 1223
 DB 914 GKQFIEKIPIRKIDFDQDEVDKYNVTWTVTEKLIITTDRIKSESNGPRRRLRLRDLAL 973
 QY 1224 -----FIASLYSLAKEDFITLLGD 1243
 DB 974 SNQLIQVINELYNISDEEYTVTLND 998
 RESULT 13
 ADC00979
 ID ADC00979 standard; protein; 1644 AA.
 XX
 XX ADC00979;
 XX AC
 XX 04-DEC-2003 (first entry)
 XX
 XX Enterohaemorrhagic E. coli 0157:H7-specific protein SEQ ID NO: 1023.
 XX

KW enterohaemorrhagic; anti-bacterial.
XX Escherichia coli; 0157:H7.
XX JP2002355074-A.
XX 10-DEC-2002.
XX 24-JAN-2002; 2002JP-00015959.
XX 24-JAN-2001; 2001JP-00112010.
XX (UYTS-) UNIV TSUKUBA.
XX WPI; 2003-451640/43.
XX Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
XX Claim 3; SEQ ID NO 1023; 2067pp; Japanese.
XX The invention relates to a novel enterohaemorrhagic Escherichia coli
CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC has anti-bacterial activity. The polypeptide can be used in detection
CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
CC genome of Enterohaemorrhagic E coli O157:H7 was determined. The present
CC sequence represents an E. coli O157:H7-specific polypeptide of the
CC invention.
XX
SQ Sequence 1644 AA;

Query Match 5.7%; Score 371.5; DB 7; Length 1644;
Best Local Similarity 17.7%; Pred. No. 1.2e-12;
Matches 246; Conservative 231; Mismatches 470; Indels 443; Gaps 51;

QY 74 QKAGQIEGVSVDILLFENDKDKASFNKKLKEAKKNEPIEDIL---IIAEVKRPTF 130
DB 143 QEEGTDLPLALLTAQFPADTDKHKHDSLRKKA--NGEYRSWQDILLSTAVFTQNEPPRF 200
QY 131 SFDAKDKLSESDQLYRYLNOYKHGYGILNGKWRLYDKSVLYGKRYIEFNFKIEE 190
DB 201 V-----LLGNRLQLLLDRTK--WAQNRLLRDFDFEILS 232
QY 191 KEEYKEQEFWFLYLIRKERYLKTNSVIEVEKEQIAKEKEI--IQKTLKEIILYERPDSS 248
DB 233 R---REYDTTLKATAVLLHKDSLPLGSGAPYLDSDNSHKHAFGVSDDLTAURE----- 284
QY 249 IVFKIAKNIYDKPKLKGKEITQHIL-----ASTLESIIFILRIFFIAY 293
DB 285 -----SIELAGNEAMHYLIDRGLANYTGNRAVDPELSRECLAYMYRLFLFY 332
QY 294 IENDIFPKILOENKLYRSSISFY-----PFYDENTKKLEYKKIITIFNLLDKGS 345
DB 333 IEARPELGAPYAKTYLQGYSLFTRDLEMIPLTSEEDRNGRYFFHDSLNMFLKLVREG- 391
QY 346 DAIKFPVNGG-----LFSDEKVKYLNGLLSISEI-----EILVKMLRPEEK 390
DB 392 -----YNGGVKMQSDLESGRITIIHSQ--FSVPRLESHLPDANNRILNRVFRNET 442
QY 391 -----NIKDEKFKYKSLRDPKSGELYETLLELYDLRIADTTVHRI----- 430
DB 443 LQIIQAMSLSRPAKGFNRGRISYRQLGINQLGAVYEALLSYRGFFASEDLVEYKKG 502
QY 431 -----IDGVVLIPT-----EELNKKVKNKIATYKGNLYLTSRLDKKSGAYTTPD 479
DB 503 EEFNELETGYFVSKDEISKYHDEKVEKDGSLRIHRKGSFYIRMAGRDKREKSASYTPE 562
QY 480 DLTDFWISSTEE-----QLKTKSPDLKIIDNSCGSHFLISCLDYLTEKVV 527
DB 563 VLTRSLVKYALKELFKQIDIPDIPSHAKADAILMLTVCEPAMGSAALFNAINQLABAYL 622
QY 528 YELDKFEDVKKELDEEYRVIIIESEYDVQDSISKELV-LKRMLLKXCIYGVDPINPISVE 586

DB 623 FH-----KQQAEGRRI-----PDQRYTQELQVRKMYIADNNVFGVDLNPVAVE 665
QY 587 ITMLSILWINTFTFGTPLSFIEHHIKAGNALLG-----YTKDEFF----- 625
DB 666 LAEVSMLNNAISGDAFVFWPGYQLHCGNSLVGARQVFNKSELTYKAKADPSWLNSEPYE 725
QY 626 -----DIVKKKPSGFSLSFKKRIKEI-----ITIL 650
DB 726 LAMNTPREBTOIFHFLPDGGMANYSDKTQRYDDPKALDSWRKEFIKSFAGHIAADV 785
QY 651 EDIYOKIKGINDTTKEDIE-----KSKKIYKEVESKD 683
DB 786 QRISEKVEALMNTYRQLKAERLTADSYVPMVAENSEQRTSLSKDETFSGRLDNSA 845
QY 684 IDNRAIIFSL-IKLYSLSPDKSLNMEFSDIASVISLIENILGNKTSSEDEKIEK----- 737
DB 846 YOKLRWMDYWCALWFWPIDKA--DELPRGTWLFIEITLLDGIIVITEKTEVAEHTTGD 903
QY 738 -----IRKLSYV-----KFFHYGIEPPDI---QE 759
DB 904 LFAEGLLREESSLSFVAGRLKTEVLFRHLPRLAIVDALRQKRPFFHMDLEFCDLFAERG 963
QY 760 GFDIVGNPPWETKXNETEFSSKHPNY--RKLGIKEQNIIOEILSKDNHPLSEYNE 817
DB 964 GFDMLGNPPWLKVEWQEAAGVLDGYEPEFVLRKLSASKLATLRIDTF----- 1010
QY 818 EKNSIIANNIYKDFK-C-----FTSGGDPNLFYRYPVTNFKLKEKGNL 862
DB 1011 --NQIPALEAAWRSEYEGCEGMQNFNAQONTYPVLGVQTNLYKCELPQAWRLGAOKGVA 1068
QY 863 TYLVPSAIWNSSSRILRKHIIFARYKLYIYOFENKRF---KDVHSSPKFAIF----- 913
DB 1069 GFHLPEGIYDDPKGQLRAAVYPRLAH--FQFQNELNLFVEVDHHAKEFSSNIYASPS 1126
QY 914 ----QLSNI-----KESTSSFKAKFWIQSS-----DNILKEITRD 944
DB 1127 VGFEHISNLYAPQTTIDACFEHSGSGDIPGLKDEIESEGLKVMVWNTSGHRSLSIATHE 1186
QY 945 LK-----DSK-DDAVKG-----TELNIQIKLSPIQESIIIEFKDNEFTL 984
DB 1187 LELFARLYDSEGTAPQARLPALHAKQLVALEKFAQNPRLGDLO-----GQVFST 1238
QY 985 INKMSKFSALGEGYI-----DFKGLDPSIKNRKSLKCEKNKNLI 1026
DB 1239 V--MFEETVTAQRDGTILROTQFPDSSQWVLSGPHFVG--TPYKTPRENTLNSVDCL 1295
QY 1027 FLYSGANIHFQNSRPPEKDAKESK---LWID--KEDEKVLTKDNQYQTERVYRAI 1081
DB 1296 DLLTLRDDYLPRTNYPACDAQEYAKRTPCVTWTELADEPKVT--DYV---RLAIRAM 1350
QY 1082 ASNTNERTWISTLSPG-----NCYCVNSIYNDEKTPISLYKGLFIISFNSFVDF 1133
DB 1351 LAQSGERTLISAIYPPPEISHMNAVRSYSSQNLLLEHGMCF-----SLPFD 1399
QY 1134 LLRRFVDSNV 1143
DB 1400 ICKSTCKANL 1409

RESULT 14
ADN46345
ID ADN46345 standard; protein; 1282 AA.
XX
AC ADN46345;
XX
DT 01-JUL-2004 (first entry)
XX
DE Thermococcus kodakaraensis KOD1 protein sequence SeqID223.
KW gene disruption; gene targeting; marker gene; transformation;
KW homologous recombination; hyperthermostable archaeobacterium; KOD1;
KW gene structure; gene function; enzyme activity; medicine;
KW forensic science; food; drug inspection; molecular biology; immunology.

[illegible]

RESULT 15
 ID AAU48711
 AC AAU48711 standard; protein; 1629 AA.
 XX
 XX AAU48711;
 XX
 XX 27-FEB-2002 (first entry)
 XX
 XX Propionibacterium acnes immunogenic protein #9607.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 KW
 KW
 OS Propionibacterium acnes.
 XX
 XX WO200181581-A2.
 PN
 XX
 XX 01-NOV-2001.
 PD
 XX
 XX 20-APR-2001; 2001WO-US012865.
 PF
 XX
 XX 21-JUN-2000; 2000US-0199047P.
 PR 02-JUN-2000; 2000US-0208841P.
 PR 07-JUL-2000; 2000US-0216747P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 PI
 XX WPI; 2001-616774/71.
 DR
 XX N-PSDB; AAS59543.
 DR
 XX
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 PT
 XX
 XX Example 1; SEQ ID NO 9906; 1069pp; English.
 PS
 XX
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: the sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX
 SQ Sequence 1629 AA;
 Query Match 5.6%; Score 364; DB 4; Length 1629;
 Best Local Similarity 16.5%; Pred. No. 3.2e-12;
 Matches 242; Conservative 246; Mismatches 531; Indels 446; Gaps 52;
 Qy 70 SVFOQKAGQTEGVESVDILLFENDKDKAFNNKLEAKK--NNEPIP-----IEDILI 121
 Db 86 SVPMPSDAILVVEDWIESEHFFTTDARKESFQKLVLDREKQWDGEDVPTTRTFTERASK 145
 Qy 122 IAEVKRTFSFADKDLK-----ESDOLYRYLYNOYKHGYILNSGKV----- 164

Db 1160 VHDLVTVGVHVGYS--PAQPHFLOASALYHPDVTVUGSLRHDGGGAPGPKVDGHWDRPH 1217
Qy 960 INQIKKLSPIQESIIEFKD-----NEEFTLINKMFSKFSALGEGYID 1001
Db 1218 AQRITETVT--DETLATWRDILDPNLEAPRTRMLYTVNRDVAETLQOLSKAPRLGSLSLR 1275
Qy 1002 FKKGLDPSIKNRKSLKXE-----CNNKNLIFYSGANIHFNSRPFEDKDAKESKLLW-- 1055
Db 1276 FSPGWHEKNORTKGYFTQQWGTDPDSWNDVIL-QGPHLH-VATPFYKSPNPTMKHNDQWSV 1333
Qy 1056 IDKEDLEKVLTKONQYOTE-----RVFYRAIASNTNERTMI-- 1091
Db 1334 VDLETLPDPAIPVTSYKPGADRAWYDADYTHWDGDPARDHYRLAWRAMAANTGERTLIPA 1393
Qy 1092 ----STLSPGNCYCVNSIYINDEKTPISLYKKLFIISIFNSFVDFLLRRFVDSNVLKSC 1147
Db 1394 IIPPGTAHPNGVFCVCGA--DNRILTA-----CAGFASLLLDIFSARAAPKSGIYQAV 1444
Qy 1148 LYQCPMP---QPEKEILSNPLYLN 1169
Db 1445 FDRLPAPCORHPLLPALLRLRLN 1469

Search completed: January 24, 2006, 19:44:49
Job time : 299.376 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 48.4458 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-16
Perfect score: 6538
Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIKGYNLLNNKIFVHK 1277

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6534	99.9	1277	2 E70224	hypothetical prote
2	5599	85.6	1278	2 B70236	hypothetical prote
3	1781	27.2	1250	2 B81339	probable restricti
4	1648.5	25.2	422	2 D70251	hypothetical prote
5	558.5	8.5	706	2 T08313	conserved hypothet
6	526.5	8.1	1181	2 B64516	hypothetical prote
7	473	7.2	134	2 E70251	hypothetical prote
8	455.5	7.0	1252	2 D71810	probable type II D
9	446	6.8	867	2 B81419	probable type IIS
10	419.5	6.4	1279	2 E64709	type IIS restricti
11	378.5	5.8	2401	2 T28676	rhostry protein -
12	371.5	5.7	1644	2 F91286	hypothetical prote
13	371.5	5.7	1644	2 B85128	hypothetical prote
14	361.5	5.5	1339	2 H81307	restriction modifi
15	359	5.5	2166	2 G70163	hypothetical prote
16	357	5.5	2269	2 T28677	rhostry protein -
17	349.5	5.3	1979	2 C71622	hypothetical prote
18	342	5.2	4981	2 T18489	hypothetical prote
19	336.5	5.1	2819	2 A90551	conserved hypothet
20	336.5	5.1	3394	2 T18501	hypothetical prote
21	336.5	5.1	3973	2 B71612	hypothetical prote
22	335.5	5.1	1939	2 T18372	repeat organellar
23	333.5	5.1	993	2 A97334	type II restrictio
24	333.5	5.1	1127	2 T28317	ORF MSV156 hypothe
25	330	5.0	2136	2 A05037	hypothetical prote
26	328.5	5.0	3724	2 T18427	hypothetical prote
27	326.5	5.0	2295	2 B71621	probable membrane
28	320.5	4.9	1650	2 T18444	hypothetical prote
29	316	4.8	1639	2 S05603	major metazoite su

30	313.5	4.8	1183	2 F90559	conserved hypothet
31	312	4.8	1516	2 E71619	RAD2 endonuclease
32	308	4.7	1631	1 SAZQK1	major metazoite su
33	308	4.7	1640	2 A24594	probable major sur
34	307	4.7	2485	1 H71621	serine/threonine-s
35	306	4.7	2829	2 A42771	reticulocyte-bindi
36	305.5	4.7	1787	2 G97222	hypothetical prote
37	305	4.7	3119	2 T18414	protein g377 - mal
38	302.5	4.6	1619	2 T18499	hypothetical prote
39	301.5	4.6	1163	2 G97236	ATPase involved in
40	301	4.6	1701	2 A26868	major metazoite su
41	298	4.6	1018	2 E64451	type I restriction
42	298	4.6	2013	2 C71610	probable membrane
43	297.5	4.6	609	2 D75633	conserved hypothet
44	297.5	4.6	971	2 A70179	exodeoxyribonuclea
45	295.5	4.5	1875	2 S38173	myosin-like protei

ALIGNMENTS

RESULT 1

E70224
hypothetical protein BBE02 - Lyme disease spirochete plasmid E/lp25
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: E70224
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A>Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: E70224
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1277 <KLE>
A:Cross-references: UNIPROT:O50698; UNIPARC:UPI0000056833; GB:AE000785; NID:g2689951; PI
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match	99.9%;	Score 6534;	DB 2;	Length 1277;
Best Local Similarity	100.0%;	Pred. No. 1.8e-241;		
Matches 1277;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKTNDIVKTNPNISLYKQLSKDFIKKENINKLKDFILIKNKLFSDDNSTEANI	ESLL	60
Db	1	MKTNDIVKTNPNISLYKQLSKDFIKKENINKLKDFILIKNKLFSDDNSTEANI	ESLL	60
Qy	61	KYIFEELNYSVEQKAGQIEGVESRDVILLFPENDKDKASFNNKLKEAKKNEPI	PIEDIL	120
Db	61	KYIFEELNYSVEQKAGQIEGVESRDVILLFPENDKDKASFNNKLKEAKKNEPI	PIEDIL	120
Qy	121	IIAEVRPTFSFPAKDLKESBQLYRYLNQYQKHGYILSNGKVRWLYDKSKVLYG	EKRY	180
Db	121	IIAEVRPTFSFPAKDLKESBQLYRYLNQYQKHGYILSNGKVRWLYDKSKVLYG	EKRY	180
Qy	181	IEFNFXXKIEEKEEYKQEWFLFYILRKERYLKTNSVIEVEKEQIAKEKEIIQK	TLKEI	240
Db	181	IEFNFXXKIEEKEEYKQEWFLFYILRKERYLKTNSVIEVEKEQIAKEKEIIQK	TLKEI	240
Qy	241	LYERPDDSDIVFKIAKNIVDKFPLSGKEITQHILASILEESIIFILRIFATYED	NDIF	300
Db	241	LYERPDDSDIVFKIAKNIVDKFPLSGKEITQHILASILEESIIFILRIFATYED	NDIF	300
Qy	301	KKTLQENKLYRSSISFRFFYDENTKKLEYKKIITIFNLDDKGSDAIKPVPFN	GGLFSE	360
Db	301	KKTLQENKLYRSSISFRFFYDENTKKLEYKKIITIFNLDDKGSDAIKPVPFN	GGLFSE	360
Qy	361	DKVKYLNNEGLLSISEETLVKMLFPFEKNIKDEKFKVYSRLDPKSGELIY	ETLLEIDL	420

361 DKVKYLNBEGLLSISBIEBILVKMLPFEBEKNIKDEKFKVYSRLDPKSGELVETLLEYDL 420
 QY 421 RIADTTVHRI IEDGVVLTIRTEBELENKVNKIATYLGNIYLTSLDRKSGAYYTPDD 480
 Db 421 RIADTTVHRI IEDGVVLTIRTEBELENKVNKIATYLGNIYLTSLDRKSGAYYTPDD 480
 QY 481 LTFDFWISSIEBQLTKSPDLKIIDNSCGSGHFLISCLDYLTKVWYELDKFEDVKREL 540
 Db 481 LTFDFWISSIEBQLTKSPDLKIIDNSCGSGHFLISCLDYLTKVWYELDKFEDVKREL 540
 QY 541 DBEYRVIISEBEYDVQDSISKELVLRMLLKXCIYGVDPINPISVEITWLSLWINTFI 600
 Db 541 DBEYRVIISEBEYDVQDSISKELVLRMLLKXCIYGVDPINPISVEITWLSLWINTFI 600
 QY 601 TPLSFIEHHIKAGNALLGYTKDEFDVKKKPESGSLFKPKIKIITILEDIYOKIKGI 660
 Db 601 TPLSFIEHHIKAGNALLGYTKDEFDVKKKPESGSLFKPKIKIITILEDIYOKIKGI 660
 QY 661 NOTTKEDIEKSKIKIYKEBESKDIDNLRIFSLIKLYSLSPDKSLNMFSDIASVISLIE 720
 Db 661 NOTTKEDIEKSKIKIYKEBESKDIDNLRIFSLIKLYSLSPDKSLNMFSDIASVISLIE 720
 QY 721 NILGNKTSSEDEKIEKIRKLSYKFFHYGIEFPDIQBGFDIVIGNPPWEKTKFNETEF 780
 Db 721 NILGNKTSSEDEKIEKIRKLSYKFFHYGIEFPDIQBGFDIVIGNPPWEKTKFNETEF 780
 QY 781 FSKHPNYKLGIGKEQNIIOBILSKONHPLSIEYNEEKNSIIANNIYKPKFCFTSGG 840
 Db 781 FSKHPNYKLGIGKEQNIIOBILSKONHPLSIEYNEEKNSIIANNIYKPKFCFTSGG 840
 QY 841 DPNLFRYFTFNLKLTKEGNTYLVPSAIWNSSSRILRKHI FARYKLNLYIQFENKQR 900
 Db 841 DPNLFRYFTFNLKLTKEGNTYLVPSAIWNSSSRILRKHI FARYKLNLYIQFENKQR 900
 QY 901 FKDVHSSFKFAIFQLSNIKESTSSFKAKPMI QSSDNILKEITRDLKSKODAYKGIELNI 960
 Db 901 FKDVHSSFKFAIFQLSNIKESTSSFKAKPMI QSSDNILKEITRDLKSKODAYKGIELNI 960
 QY 961 NOIKKLSPTQESIEIEFKDNEEFTLINKMFSKSGALGEGYIDFKGLDPSIKNRKSLKEC 1020
 Db 961 NOIKKLSPTQESIEIEFKDNEEFTLINKMFSKSGALGEGYIDFKGLDPSIKNRKSLKEC 1020
 QY 1021 NNKQLIFLYSGANIHFONSRFPEDKAKESSKLLWIDKEDLEKVLTKONQYOTERVFYRA 1080
 Db 1021 NNKQLIFLYSGANIHFONSRFPEDKAKESSKLLWIDKEDLEKVLTKONQYOTERVFYRA 1080
 QY 1081 IASNTNERTMISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISIPNSVPFDFLLRRFVD 1140
 Db 1081 IASNTNERTMISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISIPNSVPFDFLLRRFVD 1140
 QY 1141 SNVLKSLCQCPMPQPEEKELISNPLYNLAKNTSLIIAKNDPENPKYLLYLEYKFKDKE 1200
 Db 1141 SNVLKSLCQCPMPQPEEKELISNPLYNLAKNTSLIIAKNDPENPKYLLYLEYKFKDKE 1200
 QY 1201 KVNKILKLKDEBEFFKEKENNNFIISLYSLAKEDFITLLGDFAKNNKKGEDYISSL 1260
 Db 1201 KVNKILKLKDEBEFFKEKENNNFIISLYSLAKEDFITLLGDFAKNNKKGEDYISSL 1260
 QY 1261 IKGYDNYLLNNKIFYHK 1277
 Db 1261 IKGYDNYLLNNKIFYHK 1277

RESULT 2
 B70236
 hypothetical protein BBH09 - Lyme disease spirochete plasmid H/1p28-3
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
 C:Accession: B70236
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kierlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
 i Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: B70236
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1278 <KLE>
 A:Cross-references: UNIPROT:O50667; UNIPARC:UPI0000056814; GB:AE000784; NID:g2690041; P1
 A:Experimental source: strain B31
 C:Genetics:
 A:Genome: plasmid

Query Match 85.6%; Score 5599; DB 2; Length 1278;
 Best Local Similarity 85.3%; Pred. No. 6.6e-206;
 Matches 1082; Conservative 94; Mismatches 89; Indels 4; Gaps 3;

QY 6 IVKNNPNISLYKQSLKDFIKENINKNLKPFIILIKNLFSDIDNSTANIESLLKYIFE 65
 Db 12 IIKTDNPVSLYKELSGFIKENIVKSNKFFILKNKIQIADDNSTANIESLLKSIIFE 71
 QY 66 ELNYSVEQOKAGQIEGVESRVDILLFENDKDKASFNKNLKEAKKNNEPIEDILIIAEV 125
 Db 72 ELAYSVEQOKGGQIEGVKSRVDILLFENDKDKVAFNKLBEAKKNNESIPTEDILLIAEV 131
 QY 126 KRPTFSFPAKDKLXESDQLYRYLNQYKHVIGILSNKVMWRLYDKSKVLYGKRYIEFNF 185
 Db 132 KRPSFPAKDKVKEAEDQLYRYLNQYKHVIGILSNKVMWRLYDKSKVLYGKRYIEFDF 191
 QY 186 XKIBEKSEYKEQWPFVLYILIRKERYLKTNSVIEVEKEQIAKEKIIQKTLKEILYRPP 245
 Db 192 SKIKESEYKEQWFIILFYLIRKERYLKTNSIISVEKEQISKEKIIQKTLREILYRPP 251
 QY 246 DDSIVFKIANNIDKEFKLSCKEITHTLASILSEIIFILRIEFLIAVIENDIIFKKILQ 305
 Db 252 DDSIVFKIANNIDKEFKVSDKEITRHLASILSEIIFILRIEFLIAVIENDIIFKKILE 311
 QY 306 ENKLYRSISPRYPFYDENTKKLEYKKIITIFNLLDKGSDAIKPPVFNGLJFSBDKVY 365
 Db 312 ENKLYRSVSFRYPFYDENTKKLGYKKIITIFNLLDKGSDAIKFFIFNGLJFAQDKVY 371
 QY 366 LNNEGLSISIEEILVKMLPFEBEKNIKDEKFKVYSRLDPKSGELVETLLEYDRIADT 425
 Db 372 LNNEGLSISIEEILVKILPFEBEKNIKDKFKVYSRLDPKSGELVETLLEYDRIADT 431
 QY 426 TVHRIEDGVVLTIRTEBELENKVNKIATYLGNIYLTSLDRKSGAYYTPDDLTDFM 485
 Db 432 TVHRIEDGVVLTIRTEBELENKVNKIATYLGNIYLTSLDRKSGAYYTPDDLTDFM 491
 QY 486 VIISIEEQKTKSPDLKIIDNSCGSGHFLISCLDYLTKVWYELDKFEDVKELDEEYR 545
 Db 492 VIISIEEQKTKSPDLKIIDNSCGSGHFLISCLDYLTKVWYELDKFEDVKELDEYK 551
 QY 546 VIIIESEBYDVQDSISKELVLRMLLKXCIYGVDPINPISVEITWLSLWINTFI 605
 Db 552 IILKESBEYDQDSISKELVLRMLLKXCIYGVDPINPISVEITWLSLWINTFI 611
 QY 606 IEHHIKAGNALLGYTKDEFDVKKKPESGSLFKPKIKIITILEDIYOKIKGINDTTK 665
 Db 612 IEHHIKAGNALLGYTKDEFDVIKAKKPSGSLFKPKIKIITILEDIYOKIKGINDTTK 671
 QY 666 EDIEKSKIKIYKEBESKDIDNLRIFSLIKLYSLSPDKSLNMFSDIASVISLILGN 725
 Db 672 EDIERSKIKIYKEBESKDIDNLRIFSLIKLYSLSPDKSLNMFSDIAVISLILGN 731
 QY 726 KTSSDEKIEKIRKLSYKFFHYGIEFPDIQBGFDIVIGNPPWEKTKFNETEFPSKHI 785
 Db 732 KISSDEKIEKIRKLSHYKFFHYGIEFPDIQBGFDIVIGNPPWEKTKFNETEFPSKHI 791
 QY 786 PNYRKLGIKEQNIIOBILSKONHPLSIEYNEEKNSIIANNIYKPKFCFTSGGPNLF 845
 Db 792 PSYRKLSIEQNKIKQETILGRDNPNTIYNEEKNMGTINNNLYKSDPFNFASGGDPNLF 851
 QY 846 RYFVTFNLKIKEGNLTLYVPSAIWNSSSRILRKHI FARYKLNLYIQFENKGRFKDVH 905

Db 852 RYFAFNLKIKENGLTYLPSALWSSSARILRKYIFTNKGNYIYQFQNRKFDVA 911
Qy 906 SSFKAIFQLSNIKESTSFRAKFMIOSSDNILKEITRDLKSDKDDAYKGLNINQIKK 965
Db 912 TLFKAIFQISWTKTPTSFRKAFMIOSSDNILKEITRDLKSDKDDAYKGLNINQIKK 971
Qy 966 LSPQIESIEFKDNKEFTLINOMPSKPSALGBGYIDFKKGLDPSIKNRKSLKKECNQVL 1025
Db 972 LSPQIESIEFKDNKEFTLINOMPSKPSALGBGYIDFKKGLDPSIKNRKSLKKECNQVL 1029
Qy 1026 IFLYSGANIHOPNSRFFEDKAKESKLLWIDKEDLVKDKONQYQTERFYRAIASNT 1085
Db 1030 IFLYSGANIHOPNSRFFEDKAKESKLLWIDKEDLVKDKONQYQTERFYRAIASNT 1089
Qy 1086 NERTMISTLSPGNCYCVANSIYINDKTPISLYKCLFIISFNSFVDFLLRRF-VDSNVL 1144
Db 1090 DIRTWISTLSPGNCYCVANSIYINDKTPISLYKCLFIISFNSFVDFLLRRF-VDSNVL 1149
Qy 1145 KSLYQCPMPQPEEKESKLLWIDKEDLVKDKONQYQTERFYRAIASNT 1204
Db 1150 KSLYQCPMPQPEEKESKLLWIDKEDLVKDKONQYQTERFYRAIASNT 1209
Qy 1205 ILKLDKEBEPKKEKNNFIASLYSLAKEDPITLLGDFKALNKKKGEDYISSLIKGY 1264
Db 1210 ILKLDKEBEPKKEKNNFIASLYSLAKEDPITLLGDFKALNKKKGEDYISSLIKGY 1268
Qy 1265 DNYLNNKI 1273
Db 1269 ENYLNK 1277

RESULT 3
E81339
probable restriction/modification enzyme Cj0690c [imported] - Campylobacter jejuni (str
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: E81339
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Baeham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: E81339
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1250 <PAR>
A:Cross-references: UNIPROT:Q9PPL7; UNIPARC:UPI00000C1CEE; GB:AL139076; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0690c

Query Match 27.2%; Score 1781; DB 2; Length 1250;
Best Local Similarity 34.7%; Pred. No. 1.2e-60;
Matches 457; Conservative 244; Mismatches 479; Indels 138; Gaps 40;

Qy 5 DIVKTNPNISLYKOLSKDFIKENINLKDPPF-----ILIKNKLFSIDDDNSTEAN 55
Db 6 EUKYNNLNFYSLEIDFPKLYNFNANQAFRDLITKTYDKNPKIQNEHQFE-- 63

Qy 56 IESLLKTYPEELNYSVEQKAGIEGVSERVDILLFENDKOKASFNNKLKEAKQNERIP 115
Db 64 -DEFISKVLEILGWCFVRQDEKIIQKLEKDPFLFNSNDKLKSKYENLDKTKSS- 118

Qy 116 IEDILLIARVKTPTSFDKOKLSEED-OLRYLNQVQKHYGIIISNGKWLRYDKSKVL 174
Db 119 -NDFTIILES KAYNIEIDNK-KVKNPFPQILRYLGNLKNKYGFLTNGRFRFYDNS-IL 175

Qy 175 YGEKRYIBFNFKYIEKEKYEQEWFLVIYILIRKERYLKTNSVLEVEKEQIAKEKIIQ 234
Db 176 NSKNVFEINLEKIIEDQNI---EAFAYFYSVAFNFTKEKEDHLEITLQNNKLSKIE 232

Qy 235 KTLKEILVERPDDSVIFK-IANKYIDKEPKLSGKEITQHILASILEESIIFILRIFFIAY 293

RESULT 4

D70251
hypothetical protein BBK02 - Lyme disease spirochete plasmid K/lp36
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
R:Accession: D70251
R:Praser, C.M.; Caejens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: D70251
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-422 <KLE>
A:Cross-references: UNIPROT:O50806; UNIPARC:UPI00000568A0; GB:AE000788; NID:g2690123; PI
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid

Query Match 25.2%; Score 1648.5; DB 2; Length 422;
Best Local Similarity 75.7%; Pred. No. 3.7e-56;
Matches 315; Conservative 49; Mismatches 49; Indels 3; Gaps 2;

QY 853 LKLIKEGNLTLYLPSAINWESSRILRKHIIFARYKLNIIYQFENKRPKDVHSSFKFAI 912
DB 1 MKLVKPGNLTLYLTPSSLSWSESGSKALRQHIIFSNKLNIIYQFENKRGFRAMTPGKFAI 60

QY 913 FQLSNIKESTSSFKAFMIQSSDNILKEITRLDKSDKDAYKGIENINQIKLSPIQES 972
DB 61 FQISNSKEPTFRVKFVIQSSDNIMKEITSDLKEGNEYAKGIELDIAIKRLSPIQES 120

QY 973 IIEFDKNEFTLNKMFSPFSGALGEGYIDFKGLDPSIKRKSLLKCEKNKLNLIIFYSGA 1032
DB 121 IIEFRDSAEFTLVNKFSPFDTLIQYIDFREG--NLTKYKALYKEYNNEKFIIFYSGA 178

QY 1033 NIHQFSRPFEDKAKESKLLWIDKEDLEKVLTKDNOYQTRVFRAJASNTNERTWIS 1092
DB 179 NIHQFSRPFEDKAKESKLLWIDKEDLEKVLTKDNOYQTRVFRAJASNTNERTWIS 238

QY 1093 TILSPGNCYCNVIYINDEKTPISLYKKLFIISFNFSFVDFLLRRFVDSNVLSKLYQCP 1152
DB 239 TILSPKNCYCNVIYINVEIPISLYKKLFIISFNFSFVDFIIRFVNSVLKSLYQCP 298

QY 1153 MPQPEKEIILNPLYLNIAKNTSLIAKNDPENFKYLLYLFKFDKKEKWNKILKLDKED 1212
DB 299 MPQPEDEILNLSYLTIAKNTSLIAKNDPENFKYLLYLFKFDKKEEVDKMLNLPKD 358

QY 1213 EFPKKEKENNFIIASLYSLAKEDFTLLGDFKALKNKKKKGDEYISLLIKGYDNYL 1268
DB 359 EFPKKEKENNFIVASLYSLTKEDFTLLNDFKVCNK-KKGEDYISLLIKGYENYL 413

RESULT 5
T08313
conserved hypothetical protein H1130 - Halobacterium sp. (strain NRC-1) plasmid pNRC100
C:Species: Halobacterium sp.
A:Variety: strain NRC-1
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08313
R:Ng, W.V.; Ciufo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall, B.;
Genome Res. 8, 1131-1141, 1998
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasmid or m
A:Reference number: Z16408; MUID:99063795; PMID:9847077
A:Accession: T08313
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-706 <DAS>
A:Cross-references: UNIPROT:O52000; UNIPARC:UPI00000631E7; EMBL:AF016485; NID:g2822278;
A:Experimental source: strain NRC-1
C:Genetics:
A:Gene: HALOSP:H1130

A:Genome: plasmid pNRC100

Query Match 8.5%; Score 558.5; DB 2; Length 706;
Best Local Similarity 24.1%; Pred. No. 1.9e-14;
Matches 180; Conservative 150; Mismatches 219; Indels 199; Gaps 32;

QY 62 YIFELNYSVQQKAGQIEGVESVDIL-LPENDKDKASFNKKLEAKKNNEP-----I 114
DB 18 YLDEHLPETEWDVSEDEELQEAQVDDIMALWEREKDTAPKRNESQLEBEKFIKRPKRLGI 77

QY 115 PIEDILIAE-----VKRPTFSF-----DAKDKLKESED----- 143
DB 78 PFE-----VEESTSRTPRDYGFDTEDAAREREEGGDFYENAVAVADAKRWGRPL 133

QY 144 -----QLVRYLNQYKHGILSGKVMRYLYDKSKVLYGE-----KRYI 181
DB 134 DTRSGEHERDFENPSQIHVYLOETPARWAVLTDGKKRWLY-----YGTSHRLDSYY 187

QY 182 EFNFKIEEKEEYKEQEWFLVIYLIIRKERYLKTNSVIEVEKEQIAKEKIIQKTLKEIL 241
DB 188 EVDLPTILESGLDEDFKYFLF---FRHEAFLLDGGG-DCFLDDVYDESINVPAQELGEDL 243

QY 242 YERDDSIIVKIAKNIYDKFELSGKEITQHLASILEESIIFILRIFIAVIENDIFK 301
DB 244 QDN-----IYEAIKILSEGYLQYPENDIDEGDLIHDSSLIYLYFLVLYAEAG--R 296

QY 302 KILQ-ENKLYRSSISF-----RYFFYDENTKKLEYKKIITIFNLDKG 344
DB 297 ELLDTSNIYEQSYLSNLSLKQEIABELDSGPKYRDMQDNIQSRLD-----ELFTLINK 351

QY 345 S-----DAIKFPVFNGLFSED-----KVLYLANE--GLLSISIEIEILVKMLPFE 388
DB 352 SKSRGIPEDLYIIPAYNGGLFRTDPDEDDSAEAKFLANHDVGDAYLAKVVELLTR---S 407

QY 389 EKNIKDEK-FVKYSLRDPKSGELYETLLELYDLRIADTVTHRIIDGYLIRTBELNKK 447
DB 408 KNDGGGKIFVDYSSLDVRHLGSIYEGLLYQLANVADEPL--ALDDGYV--SADEGDDV 463

QY 448 KVKIATYKLNIVYLSRSLDRKSGAYYTPDDLTDNV-----ISSIEELQATKSP- 499
DB 464 VVQ-----EGEVLTDTDSGERKATGSYITPEYVVEIVENTLEPLVDDIDRADLAGRSAR 517

QY 500 -----LDIKIIDNSCGSHFLISCLDYLTEKVMYELDKFEDVKKELDEBYR 545
DB 518 GEDRGFAFEAFERIFDLKILDPMGSGHFLSAIDYLAREI-----IDAQER 564

QY 546 VRIIE-----SEEDVDQDSISKELVLRMLKXCIYGVNDINPISVEITWLSWINTFI 600
DB 565 QAAQQGIETVQDHDINWA-----RQVAQRCIYGVDLNPLAVELAKVLSLWLTAAE 617

QY 601 TPLSFIEHHIKAGNALLGYTKDEPFDIVKKK-----PESGFSLPKRIKEIT 648
DB 618 QPLAFDLHLTKTGNLSVGSNVEE-IDALERNGAABEGPNSTLIDFGMT-FEGTWNLMWA 675

QY 649 ILEDIYQKIKGINDTTKEDIEKSKYIK 676
DB 676 IHADLI-----ALENEELEDVKAMEKKYQ 699

RESULT 6
B64516
hypothetical protein MJEC502 - Methanococcus jannaschii plasmid pURB801
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: B64516

Db 62 EDLEKSKYIYEQENINNLRIIFSLIKLYSLFSDKSLNIKFSDITVINLIGNILDS 121
 726 KTSSDEKKEK 737
 122 KISSEDEKKEK 133

RESULT 8
 D71810
 probable type II DNA modification enzyme (methyltransferase) - Helicobacter pylori (stra
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C:Accession: D71810
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jlang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557; PMID:9923682
 A:Accession: D71810
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1252 <ARN>
 A:Cross-references: UNIPROT:Q9ZJA4; UNIPARC:UPI00000D373F; GB:AE001563; GB:AE001439; NID
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp1409

Query Match 7.0%; Score 455.5; DB 2; Length 1252;
 Best Local Similarity 20.7%; Pred. No. 3.1e-10;
 Matches 302; Conservative 202; Mismatches 499; Indels 453; Gaps 61;

Qy 22 KDFIKEN-----INLKDFILKNKLFSDDDNSTEANIESLLK--YIPELNYS 70
 Db 9 KVFIIKYNQPEPKGTIENFEKEINSLLENVPRQDDEEFQKNEINKFLKNAYGDCNTY- 67
 Qy 71 VEQKAGOLEGVESVDILLFENDKDKASFNKLEAKKNNPIP-----IEDI- 119
 Db 68 ----EKVDSVIYDKEVRVLI-----EVKALNNKTEFPKNNPLSKACQWVLYFLSEIE 119
 Qy 120 -----LIIAEKVRPTFSDAKDKLKESED-QLYRYLNOYQKHGYLSNGKWRLYD 169
 Db 120 KEKNNSLKHTIICNA-HBFFLFDCKDLSLKEDKRIKKFYKNYAKKEGT-----D 168
 Qy 170 KSKVLY-----GEKRYIEFNPXKIEEKEKQWVLFYILIRKERYLKT 216
 Db 169 SSKPKFYKDLQEQEDFQGLRYTHFNJL-----SDDFKE---LPLIYQVLSQEVLLKQK 220
 Qy 217 NVIEVEKEQIAKEKEIOKTLKEILVERPDD--SIVFKIA--KNIYDKEFKLSGKEITQH 272
 Db 221 RTLDAN-----TLNKDFYEBLLYILGLEQNDYKGLIKIPSRNLSLGAKEKYKNLQDE 276
 Qy 273 ILASILEESIIFILRIFFIAYIEDNDIFKKILOENKLYRSSISPRYF-----FYDENTK 326
 Db 277 EWMALL---IANNRIILFLRLESLLISPEHPEKERSFLTTGNCFQNALNTLFLPEVLAK 333
 Qy 327 KLEYKKIITIINLIDKGSDAIKFPVNGGLPSE-----DKVKYLANNEGLLSISEIE 379
 Db 334 KNSRLPEIKENKILG-----KIPYLSNLSLFDKPTLELKGHEIKLLDNKKL-----E 380
 Qy 380 ILVKMLFPEEKNIKDEKFKVYSRLDPKSPGELYETLLVEYDLRIAD-----TVHRIEDGV 435
 Db 381 IYNSVLKHEYNQDK-----ASLPLEYFFKFLHLYKFTTPDKDKONT 426
 Qy 436 YLIRTEEELNKKVANKIATYLNKGNITLTSRSLDRKSGAYYTPDBLTDPMVSISSIE---- 491
 Db 427 --DTSESRLINPSV-----LGLVPEKLVNGYKESGYFTSPITSYMKCESITTVL 474
 Qy 492 -----BQLKT-----KSPDLKIIDNSCGSGHPLISCLDYL 522
 Db 475 DKFNOKYNIETCDKJELKNLYLNKSNYKEDKRKYQLQLLTLCRDPVAGSGHFLVSNALNEM 534

Qy 523 TEKVVYELDKFEDVYKCKLDEBYRVIIESE-----EYDVQDSISKE 563
 Db 535 V-LIAYELGLIASLHR-----YSLTENDEIIIHVTPGEIFNYKKPLRENDPHHHIQKE 588
 Qy 564 LV-LKRMLLKKCIYGVDPINPISVEITMLSLWI-----NTFIEGTPLSPFIEH 608
 Db 589 LFEKLSIIEINCLFGVDINPNSCBITKLRLWIELLYKSYIIFEBGKNTNALET-LPNIDI 647
 Qy 609 HIKAGNALLGYTKDEPFDDIVKKPESGFSPLFKKRIKEIITIL-----EDIY 654
 Db 648 NIKCANSLI--SRNMLNDDLUK-----IPNIKKIQEYKLVIAQYKDPNPYLPNKADLI 700
 Qy 655 QKIKGINDT-----TKEDIEKS-KKIYKEYEESKDIDNLRIRIIFSLIKLYSLSPDKS 704
 Db 701 NKIQDLAKNTFSLTKDPKTKAELEKTEKHKKYN-----FPAL-----DDKS 743
 Qy 705 LNMFSDIASVISLIENILGN-KTSSDEKE-----KIEKIRKL-----SSYKFPFH 749
 Db 744 L-----LDGLNYFIPSLFGLTKLSPKEEBEAFASYGRIRALRKKLDLDDALSGBGEYQNAFE 797
 Qy 750 YGIEFPDIOE-----GFDIVIGNPPWEKTKFNETEFPFSKHIPNVRKLGIKEQNIILKOEI 803
 Db 798 WRFEFPEVLDDDEGFLGDCIIGNPPY-----IROEQ 829
 Qy 804 LSKDNHP--LSIEYNEEKNSIIAIINNIYKDFKCTSGDPPNLFRYFTFNKLKLIKEGNL 862
 Db 830 I-KDKLPLEKQYQDYFNS-----TADIYTFYFALSFPHLLEKGF 869
 Qy 863 TYLVPSAIWNESSRIILRKHFARYKLNYYIYQFNKRFKDVHSSFRPAIFOLSNIKEST 922
 Db 870 AFITSNYARAKYGAKLREMLKFTLVSYMELNALK-----VFESAADVTSI 917
 Qy 923 SSRAKAFMIQSSDNLKTEITFDLADS-KDDAYKGIELNINOIKKLSPIQESIIEPKDNEE 981
 Db 918 IHFIKQPLKESDFKYEPTDDDKDLKSTPYLSMKQNALSTESFIFANATLLDLRD--- 974
 Qy 982 FTLINKMFKSFALSALGEGYIDFKKGLDPSIK-----NRKSLLEK-----NN 1022
 Db 975 -----KIESVGTLPKWDIOINYGIKTGANEAFIIPTEKREDEILNACKTQBERERT 1025
 Qy 1023 KNLI-----FLYSGANIHOFSRPFEDKDKAKESKLLMIDKE-----D 1060
 Db 1026 ERLIKPILRGKDIKRYSYEWAGELINHTNGYTSNLK--SKIPPIDIEKYPATKAHLDAH 1083
 Qy 1061 LEKVLTK---DNOYQIYERFVYRAISNTNERTWISTLSPGNCYCVNSIYINDEKTPISL 1116
 Db 1084 YDTIATRCQDGTTPYHLRNCAY--LEDPEKEKIVYPETSSQGFYIYENSGIFLEKTVFMI 1141
 Qy 1117 ---YKKLFIISFNSFVDFLLRRFVDSNVL-----KSCLYQCMPQPEEKEILSN 1164
 Db 1142 VSDAYNLKLLTALLNSKLITFYFNFCGCGCILKSGYQYNKHALEKIPQITEKNQBELA 1201
 Qy 1165 PLYNLAKNTSLLIAKNDPENFKYLLYLEYFKPKDKVKYNKILKDKEDFEFFKKEKNENNF 1224
 Db 1202 DKITDCAEQIILAKAKOPKANTQKL-----EKEIDA 1232
 Qy 1225 IIASIYSLAKEDFITL 1240
 Db 1233 LVYQLYHLTDDEIKTI 1248

RESULT 9
 B81419
 probable type IIS restriction /modification enzyme, N-terminal half Cj0031 (imported) -
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
 C:Accession: B81419
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vanvliet, A.; Whitehead, S.; Barrell
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: B81419

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-867 <PAR>
A:Cross-references: UNIPROT:G9PJ080; UNIPARC:UPI00000C2009; GB:AL139074; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics: CJ0031

Query Match 6.8%; Score 446; DB 2; Length 867;
Best Local Similarity 22.4%; Pred. No. 4.7e-10;
Matches 228; Conservative 141; Mismatches 291; Indels 358; Gaps 45;
QY 34 KDFPILKKNKLF---SIDNSTEANTIESLLKXIFELNYSVBOQKAGQIEGVESRDILL 90
DB 9 KDFP---NPPYRKQIMQNEFIDISKALMQYL-FELSSQSENEEDYLVANALSPFLTWL 63
QY 91 FENDKDKASFNKKJEAQKNBPPIPI-----EDILIAEVKRP-TFSPDAKDKLES 141
DB 64 -----NFKTHIKTKQKSEIDLISKDFSKDLVLEIAKKPNSEKFIHTKVNK 115
QY 142 ----EOLYRYLNQYKHGYILNGKWRLYDKSVLYGKRVIEFNFXKIEKEE--YKE 196
DB 116 ALHETILYYFRN-----REYSPSLFPIITDFYKFIKISFEELFYKN 160
QY 197 QEMFVLFYILIR-----KERYLKTSNVIEVEKEQI-----AKEKE----- 231
DB 161 PSFKKLFEBFCNPNSLFGKNTTEFYKEVAKLGENSKENLKGFLIDLTLFKOKQKNFKNL 220
QY 232 -LIQKTL-----KEI-----LYERPDDSIYFKIAKNIYDK-- 261
DB 221 ASIYKTFHRDFLLSEFNPNANSLNNAFYKELLYILGYESKONSLI-IAKSEESKEQ 279
QY 262 ---FKLSGKEITQHILASILESIIFILRPIFIAYIEDNDI-----FKKILQE 306
DB 280 GTFYTAINSKLEENFETILKLLILNLRILFKLIESNLVRFNDDKNLFLNFKKIPDF 339
QY 307 NKLYRSSISFRFFDYDENTKKLYEKKIITINLLDKGSDAIKFPVNGGLSEDKVKYL 366
DB 340 DKL--SELFFVLAKEKTRKKSEFTYL-----PYNSSLFKOSI--- 378
QY 367 NNEGILSIEIEILVLMVLFEEKNIKDEKFKVYKRLDPKSGELVETLLVDLRIADTT 426
DB 379 --ENTLEISLNDL-KLFFYKNTVLKDK-CKAKGQVGLLEYLFEFLDSDFGSD-- 432
QY 427 VHRIIEDGVYLIRTEELLENKVNKIATYLNKNIYLTSLDRKSGGAYITPDDLTDFMV 486
DB 433 -----EQSEILLSQKEL--ISSSVLGNVF--EKLNGYKESGYTPSFITSYMC 475
QY 487 ISSI-----EQLTKSPL--DIKIINSCGSHFLIS 517
DB 476 KESITKVLDFKNAQFDLDAKDISLRKSLRKEDKKAQKELLNSIKICDPVAGSGHFLVS 535
QY 518 CLDYLTEKVMYELDKFEDVKKLDERYVYIESE-----EY-----DVQDS 559
DB 536 ALNVMV--SIYDENLF-----DEEFLVQNDLEILITGRKGBEIEYKRPSTPKDTHL 587
QY 560 ISKELV-LKRMLLKXCIYGVNDINPISVEITMLSLMINTPIFGTPLSF----- 605
DB 588 IQQELFHTKKDIENNLFGVDINPNSCEITKLRLMIELLKHSPYQSFDDGNYHDLKTLPN 647
QY 606 IEHHKAGNALLGYTKDEPFDIVKKFPGSGFLFKRKEIKETIILEDIYQKIKGINDTTK 665
DB 648 IDINIKCGNSLVSY-----FETGKSL-----NHPYNI-----K 675
QY 666 EDIEKSKYIYKEE-----SKDIDNRIIF-----SLIKLYSL 699
DB 676 ERMGYKRIYDKYGEFTYTKSHINOIEKNLKSIFKNFCFADKFKKEMKSFNDCKEKYSK 735
QY 700 SPDKSLNMFSDIASVISLIENILNGKT--SSEDEKIEKIRKLSYYKF-----F 748
DB 736 KYGNFLAVDDENLKPVS-----ANLTLFDFDEKAIKEPANLKEKYDNIENLSNHPF 789
QY 749 HYGIEFPDIQE-----GFDIVIGNPPWEKTKFNETEPFSKHPNYRKLGIKEQNIKQE 802

DB 790 EWRFEFPBILDDGNGFKGFDLIIGNPPY-----IRQE 821
QY 803 ILSKDNHPLSIYNEEKNSIITAINIYKDFKCFSTGSDPNLFRYVFTFNKLKIEKG 860
DB 822 -----ELKELKPHLAKNYKVKY-----GTSDIYTYFYELGFNVLDKRG 859
RESULT 10
E64709
Type IIS restriction enzyme R and M protein - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C:Accession: E64709
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, S.; J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A>Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: E64709
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1279 <TOM>
A:Cross-references: UNIPROT:O26046; UNIPARC:UPI00000D30D2; GB:AE000649; GB:AE000511; NID
Query Match 6.4%; Score 419.5; DB 2; Length 1279;
Best Local Similarity 21.1%; Pred. No. 7.5e-09;
Matches 308; Conservative 209; Mismatches 454; Indels 487; Gaps 70;
QY 68 NYSVEQKAGQIEGVESRDILLFENDKD-----KASFNKKLEAKK----- 109
DB 12 NYPNQ-----QKAPETAFDAFLETNQOENENHQNDLLKGFKYKVPKTKIDSTI 67
QY 110 --NNEPIPIEDILIAEVKRPSTFSDAKD-KLKESEDQLYRYL-----NQYQKHGYIL 159
DB 68 LNENNE-----VEVIEFKALKNPNEFIKGGDLNVKAFHESLSLYLTERKEGNNLKL-IL 123
QY 160 SNGKVMRLYDKSV-LYGEKRYIEFNFXKIEKEEYKEQEWFLFIY-----LIRKERY 212
DB 124 ATIKELYIIDANEFEVFNKDEIENAFKNCHDR---KGNDRTRKAFYDACQKRLNEFDRS 180
QY 213 LKTSNVIEVEKEQIAKEKIIQKTLKEILYERPDDSIYFKIAKNIYKFKLSGKE----- 268
DB 181 LK-VHYIPLKKNLAL--IYQALSPNELLKIPKYSNDANTLNKOFYEBLLYILGLEQND 236
QY 269 -----ITQHILASIL-----EESIIFIL-----RIFFIAYIEDNDIFKKILQE 306
DB 237 KKKILIKPSRTQNSLSDALKKEYKNLDDDEVWALLIANNRILFLRLLES----- 286
QY 307 NKLYRSSISFRFFPYDEN---TKKLEYKKIITIF--NLLDKGSDAI-----KF 350
DB 287 -----LLISFKHF--ENPFLTTFENFNDLNTLFFVLAKNSERLPEIKEDKILEKI 338
QY 351 PVFNGGLFSE-----DKVKYLNNEGILLSIEIEILVLMVLFEEKNIKDEK---FVKY 400
DB 339 PYLNSSLFDKTPLELKGHEIKLLDNKKL-----EYKNSVLKHKHQYKEKPLPLKY 391
QY 401 SRLDPKSGELVETLLEVDLRIADTTVHRIIEDGVYLIRTEELLENKVNKIATYLNKNI 460
DB 392 -----LKFRLRYKF---TTTPDKDKONT--DTSESRLINPSV-----L 425
QY 461 YLTSSLRDKKSGAYITPDDLTDFMVISIE-----EQLKT----- 496
DB 426 GLVFEKLNKYEGSYTPSFITSYMKESIPIVLDKENAIYQWDCENLKGRLGEIDRNF 485
QY 497 -----KSPLDIKIINDSCGSHFLISCLDYLTEKVMYELDKFE-----DVKKEGDE 542
DB 486 SNEKAYELNTLLTLCRCDPAVSGSHFLVSNLNMV-RVAYELGLIASLYRYDLKLENDE 544
QY 543 -----EYRVIIIESEEDVDQDSISKELV-LKRMLLKXCIYGVNDINPISVETMLSL 592

Db 545 IIIHHTPTGGEIFYNIKPDSNDPHHHIOKELFNLKKSIIENCLFGVDINPNSCEITKLRL 604
Qy 593 WT-----NTPIFGTSLPSFIEHHKAGNAL-----LGY 619
Db 605 WIELLKYSYIYEKKNNTALST-LPNIDINIKCANSLISRPALKDKKALLKSEKNLLEY 663
Qy 620 TKDEFDFIVKKFSGFSUFK-KRIKEIIT-----ILEDIYOKI----- 657
Db 664 STAEYKELVK-----IYKPKILETLTHPIKDSNAVRKYAKERLYOELQNPNKDFK 715
Qy 658 KGINDTTKEDIKSKYIK--BYESKOJDNRIIFSLIKLYSLSPDKSLN---MEFSDI 712
Db 716 KALNDR-----IEKIKAFKLTLPNPPPKELPKFKFLKEHLELYCKSLSEANTNGLEAL 771
Qy 713 ASVISILNIENLGNKTSSEKKEIKRLSSYKF-----PHYGLEF 754
Db 772 A-LEKOMANLFFDRYPKLDKSDKVVGLEHFNRYVLTSYKQLODENERYANALEWREF 830
Qy 755 POIQE-----GFDIVIGNPPWEKTKFNETEPFSKHIPNRYKLGIKEQNIHQEILSKDN 808
Db 831 PEVLDDGDFSGDCIIGNPPY-----IRQEH-KDL 861
Qy 809 HP-LSIEYNEEKNSIIANNIYKDFPKCFTSGDGNLFRYFVTNFKLKEKGNITYLVP 867
Db 862 KPILLEKOYODFYNS-----TADIYTFPALAFHLLKEKGFSAFITS 902
Qy 868 SAINNESSRIILKHIFAR-----YKLNVIYQFNKKFKPDVHVSFKFAIFQLSNIKES 921
Db 903 NKYTRAKYGAKEWELKKTTIVSYMELNALVKFESAA-----VDTSI-----IHFIKOT 952
Qy 922 TSSFRKFMIOSSDNLKEITRDLDKSDDAYKGIELNINOIKLSPLOESIEFPKDNBE 981
Db 953 PSK-----ESEFKYFETPNDKDLKSTPHLLMKQNVLTSTESPIFANATILDLRD--- 1002
Qy 982 FTLLNMFKFSALGEGYIDFKKGLDPSIKNRKSLLEKCNKNLIFLYSGANIHOFNSRF 1041
Db 1003 -----KIESVGTPDKWDIQINYGIK-----TGA-----NEAF 1030
Qy 1042 PEDKAKESKLLMWDKEDLEKVLTKDNQYQTERVPY-----RAIASNTNERTMTSLSP 1096
Db 1031 IIPTEKRE-----EILNACKTOBERERELIKPILRGDKIKRYSYEWAWHLWVINT 1081
Qy 1097 GNCYCVNSIYINDEKPIISLYKKLFIISFNSGFVDFLLRRFVDSNV---LKSCLYQCPM 1153
Db 1082 HNGYTS--LASKPIPIDIEKYPAKALHDAH-YDTIATRCQDGTPHLRNCAYL--- 1134
Qy 1154 POPEEKE-----ILSNPLYNLAK-----NTSLLI----- 1178
Db 1135 -EDFEKEKIWASVGFVEYCMIPGLLILDTNVPFEVSKFGNTKNVLLGLLNSKLLTFMLK 1193
Qy 1179 AKNDP-----ENFKY-LLYLEYFKF---DKEKNVILKLKDEDEFFKEK--- 1218
Db 1194 AKNTPLGDMGAVRNYKYNIMELPMWITAKNKXKIADKIALVDKILQAKBKDPKANTQKL 1253
Qy 1219 ENENNPIIASLSLAKED 1236
Db 1254 EXEIDALVYQLYHLTDEE 1271

RESULT 11
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
A:Accession: T28676; A45521
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>

A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; P1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple c
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 5.8%; Score 378.5; DB 2; Length 2401;
Best Local Similarity 20.0%; Pred. No. 5.6e-07;
Matches 330; Conservative 247; Mismatches 527; Indels 547; Gaps 74;

Qy 14 ISLYKQLSKDPFKKENINKLXDF-----FILLKNKLFSDDDNSTEAINESL-----LKLYFE 65
Db 253 VDLKKEIBKDNVYIDELAKEPPYQITKYIEKKEIY-----NTIKSDFDKIYVGDIEQLYN 308
Qy 66 ELNYSVEQKAGQIEGVESRVDIILLFENDKDKASPNKKLKEAKGNNEPIEDILIAEV 125
Db 309 EM-FSVVQ--SNIEHIENKTEILTAKIDNVNINQMETE-----TV 350
Qy 126 KRPTFSFDPAKDLKXESDQLRYNLQYQKHYGILNSGVWRLYD---KSKVLYGE----- 177
Db 351 KSHLKNIEIETNNKLGSETILDIKIYI-----YGEITNELNKTLEDPKNEKGLSNKIDEVA 404
Qy 178 KRYTEFNFK--TEEKEEYKEQWFLFYILIRKERYLKTNSVIEVEKEQIAKEKILQ 234
Db 405 KENVQILNYKSNILEIKKHYNQ-----INIDNIKEKEAKQNYDQFKHEM 449
Qy 235 KTL--KEILYRPPDSDIVFKIAK-----NIYKFEKLSGK-----EITOHILA 275
Db 450 KIIPNEMKYQKP--SIEIKMKDEFLSKVKNYNDFKVYKEVSEHNKFTTELNTKTKT 507
Qy 276 SILSESIIFILRFIAYIEDNIPFKILQENKLYRSSISFRYPFYDENTKKL--EYK 333
Db 508 EVSDEEI-----KK--YENKFNDSKSLI-----NETKKSIEEYQN 541
Qy 334 IITFNLLDKSDAIKFPVFNGLFS-----EDKVKYLNNEGLISIEIEILVKM 384
Db 542 I-----NTLKKVDDVIKVLNTNELITNCHNKQTTLKDKL-----NQNKITKETNSI-DKI 592
Qy 385 LFFEEKNI-----KDEKFVKYSLDPKSPGELVETLL--YDLRIADTTVHRIEIOGVYLI 438
Db 593 YTDKFENILTKKTELETFTGLSLNHNESNNKELLTYFYDK---ANLKKKENWLYKQ 649
Qy 439 RTEEB-----LENKKVNIATVILKGNITLSRSLDRKSGAYYTPDDLTDFMV----- 486
Db 650 FNEKEKAVEDIKKQVNDINKIVSNIEITITYSIYNEDTENEIGKSIELLNTKVKLEKVK 709
Qy 487 -----ISSIEBQLK-----TKSPLDIKIIDNSCGSGHPLISCLDYLTE----- 524
Db 710 ANVTNLNBIKEKLKDYDFQFKEKNIKYPDENKIKNDIDTLNQKIDKSIETLTITKQNS 769
Qy 525 -----KVVVELDK-----FEDVKKELDEEYRVITEESSEYDVQDSISKELVKRML 570
Db 770 ENHIDEIKGIDKLKKVPNKTMFNEDPKIEKKIENIV---EKIDKKNIYKEI---DKL 823
Qy 571 LKXCIYGVNDINPISVEITMLSLWIN-TFIFGTPLSFI-----EHHIKAGNALL 617
Db 824 LN-----EISKIENDKTSLEKLNINLSYKSLGNLFLQIDEEKKKAHTIKAMEA-- 875
Qy 618 GYTKDEFFDVKKPFSGFSFKRIKEIITILEDIYOKIKGIN-----DTRKED 667
Db 876 -YIDD--LDNITKKSQE-----IEKEMNINMDIKMDIHKEMKALNTHDDYKIYHTTSKNH 928
Qy 668 IEK-----SKKIYKEEESKDIOLRAIFSLIKLYSLSPDKSLNMFSDIASVLSLIE 720
Db 929 EEKISDIRKNSLKIIQDFSEESYINDIKKELEKNVLESQNNNTDINQYLSKLENIYNI-- 987
Qy 721 NILGNKTSSEDEKIEKIRKLSYKFFHYGIEFPDIOEGFPDIVIGNPPWEKTKFNETEF 780

Db 988 -----KLKIKKIIDKVK--EYTDIEIKNNKINAEIENSKLIITQIKENSS 1032

QY 781 PS-----KHIPNRYKLGKIEONIKQIEILSKDNHPLSIEYNEEKNS 821

Db 1033 LKQCSKIKSTIDDNVSECIKNITNLTKYIIVNEKNNTYFPKNAE-----EYNQ--NV 1084

QY 822 ITAINNIYKDFK-----CFTSGDGNLFRYFVTFNLKLIKKEKGNLTYLPSALWNESS 875

Db 1085 SLNFNNIEMADTKSQYILNIKKNGTNN-----TDYNIKELKEHKKGNVYKDEAGKNTQ 1139

QY 876 SRLRKHIPARYK-----LNYIYOFENKRF-----KDVHSSF----- 908

Db 1140 EIKKKNELFEKVEQVTVLKNKYAVELKKNKDKTKYNSQYIIKEIKDAHNTFTSQADKS 1199

QY 909 -----KFAIPOLSN-----IKESTSSFKAF-----MIQSSDNILKEIT 942

Db 1200 EKOMEIKNEQIRIEDEVAKNKSNKAILDIQLSVEPPKIKPLKIKDLRTKSSDCLKE--T 1258

QY 943 RDLK-----DSKDAYKGI 956

Db 1259 KDIEFKISNLSIDTQETKLIENKNILNLEKLESKKNQKNIEQKELDEVNSKIKNI 1318

QY 957 ELNINOIKK-----LSPQIESIIEFKDNE-EPTLNK 987

Db 1319 ESNVQHKQYEGIVEKINEIAKANKQIOESTQKLIPTIKNLISPPKANDLEGIDTYK 1378

QY 988 MFSKPSA-LGEGYIDFKGLD-----PSIKNRKSLKCEKNNKNI-----FLYSGANIHQ 1036

Db 1379 NLGKYNTEMNNIYEFIKSYDLITHYLETVSKEPITYEQIKNKRTAQNELLTNIKVNK 1438

QY 1037 FNSRFEDKAKESKLLWIDKEDLEKVLTK-DNQYQOTERVYPAIAS-----NTHER 1088

Db 1439 AKS-YLDDIEANEPDRIVTHFPKNKLDVNDKFTNEYSKVKNKGFDNISNNVKKSTDEN 1497

QY 1089 TWISTLSP-----GNCVCVNSIYINDEK-----TPISLYKCLFI 1122

Db 1498 LLNLINQTKENYANIVSKYYSKYEAENIFINIPKLANSLNIQIKSSGIDLPFNINI 1557

QY 1123 -----ISFNSFVDFLLR----- 1136

Db 1558 AILPYLDSKDDTLAFIPSEKTSYTYTKISDSYNTLLDLKRSQELQKQOQALNLIFE 1617

QY 1137 -----RFVDSNVLSKLYCQMPQOPEKEILSNPLYNLAKVTSLLTAKNDPENFKYLL 1190

Db 1618 NRLHDKVQATNELKDTLSL-----KNKKEQILNKVKLLHKSNELNKLSNCSQNYDTI- 1672

QY 1191 VLEYEFKDKVKYNKILKLDKDEPFKEKENEN-NFIILASLYSLAKEDFTLLGDFKALKN 1249

Db 1673 -LESSYDKIK-----EKSNNYKEKENLGINFDVRAM-----BEQFNNDIKOTEKLEN 1720

QY 1250 KKKGBDYISSLIKGYDNVLL-----NNKIFYHK 1277

Db 1721 NYKHSE-----KDNYNFSENNNLIQSK 1743

RESULT 12

F91286

hypothetical protein ECs5262 [imported] - Escherichia coli (strain O157:H7, substrain R1

C:Species: Escherichia coli

C:date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: F91286

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; PMID:11258796

A:Accession: F91286

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1644 <HAY>

A:Cross-references: UNIPROT:Q8XC72; UNIPARC:UPI00000D04EE; GB:BA000007; PIDN:BA83685.1;

A:Experimental source: strain O157:H7, substrain R1MD 0509952

C:Genetics:

A:Gene: ECs5262

Query Match 5.7%; Score 371.5; DB 2; Length 1644;

Best Local Similarity 17.7%; Pred. No. 6.7e-07;

Matches 246; Conservative 231; Mismatches 470; Indels 443; Gaps 51;

QY 74 QKAGQIEGVESEVDILLFENDKOKASFNKLEKAKKNPIPIEDIL---IAEVKRTPE 130

Db 143 QEEGTDLPLAUSLTAQFPADTDKHKDSLKKKA--NGEYRSWQDLSTAVFTQNEPRF 200

QY 131 SFDKDKLKESDQLYRYLNQYKHGILSNGKVMRLYDKSKVLYGERRYIEFNFKIEE 190

Db 201 V-----LLNGRQLLLDRTK--WAQNRLRDFDEEILS 232

QY 191 KEEYKEQWFLFYILIRKERYLTKSNVIEVEKQIAKEKEI--IQTKLKELIYRPPDS 248

Db 233 R---RETDTLKATAVLLHKDSLPCSGAPYLDSDNSHKHAFGVSEDLKVALRE---- 284

QY 249 IVPKIANNIYDKFKLSKEITQHL-----ASILEESIIFILRFFITAY 293

Db 285 -----SIELLGNEAMHYLIDRGLANYTGNRAVDPDELSRECLRYMYRLFLFY 332

QY 294 IEDNDIPKKILOENKLYRSSISFRY-----PFYDENTKKLEYKIIITIFNLLDKGS 345

Db 333 IEARPELGYAPWTAQTYLOQYSLETLDLEMIPLTSEEDRNGRYPHDSLNLMLFKLVREG- 391

QY 346 DAIRPPVFGG-----LFSEDKVKYLNNEGLLSISEIE-----EILVKMLFFBEK 390

Db 392 -----YNGGVKMQSDLESGRDITIHSQ--FSVPRLESHLFDANNTRILNRVVRNET 442

QY 391 -----NIKDEKFVKYSLDPKPSGELYETILEYDLRLTADTTVHRI---- 430

Db 443 LQOIQAMSLSRPAKRNRRGRISYRQLGINQLGAVYEALLSYRGFFASEDLVEYKAG 502

QY 431 -----IEDGVYLIRT-----EELENKKNKVIATYKNGNIYLTSLRSLDRKSKGAYTPD 479

Db 503 EEFNELEGTGVFVKDEISKYHEDEKVKESLHRKGSFTYRWAGDRKESASYTPE 562

QY 480 DLTPMVISSIEE-----QLTKSPLDIKIIDNSCGSHFLISCLDYITEKVM 527

Db 563 VLTRSLVKYALKELFKEQIDPISPHAKADAILNITVCEPAMGSAALFLNEAINQLAEAYL 622

QY 528 YELDKFEDVKELDEEYRVIIESEEDVDQDSISKELV-LKRMLLKXCIYGVDPINPISVE 586

Db 623 FH-----KQQAEGRR-I-----PODRYTOELQRVQMYIADNNVFGVDLNPVAVE 665

QY 587 ITMLSLMINTIFIGTPELSFIEHHIKAGNALLG-----YTKDEFF----- 625

Db 666 LAEVSMLNALISGDAFPWFQYQLHCGNSLVGARQVFNKSELTYTKKADPSWLNSEPVE 725

QY 626 -----DIVKKKFSFSGFLFKRIKEI-----ITIL 650

Db 726 LAMNTPREETQIFHFLLPDGGMANYSDKTQVQRYPDDEFKALDSWEKFIKSPAGHEIADV 785

QY 651 EDIYQIKINGINDTKEDIE-----KSKKIYKEVEESKD 683

Db 786 QRISKEVEALWNTRYQQLKAERLKTADSYVPWPAENSEQTRSSLSKDETFSGRLDNSA 845

QY 684 IDNRIIPSL-IKLYSLSPKSLNMFSDIASVLSIENILGNKTSSEDEKIEK----- 737

Db 846 YOKLRWVNDYWCALWFWDIDKA--DELPRGTWFEIETLDDGIVITEKVTVAEHTTGD 903

QY 738 -----IRKLSSYY-----KFFHYGIFPPDI---QE 759

Db 904 LFAEGLLRRESSLFSVAGRLKTEVLFPHLPRLAIVDALRKHQRRFFHWDLFFCDLFAERG 963

QY 760 GFDIVIGNPPWEKTKFNETEFFSKHIPPY--RKLGIKEQNIKKOILSKDNHPLSIEYNE 817

Db 964 GFDMLGNPPWPKVQEWQAGVGLGYEPFVLRLKSASKLATLRIDTF----- 1010

QY 818 EKNSIIANNIYKDFK-C-----FTSGGDPLFRYFVTFNLKLIKKEGNL 862

Db 1011 --NQIPALEAAWRSEYEGCEGMQNFNAQNYQVPLRGVQTNLYKCFLPQAWRLGAOKGVA 1068

QY	863	TYLVPSAINWESSRILRKHIIFARYKLNYYIQENKKRF---KDVHSSFKFAIF-----913
Db	1069	GFLHPGEGIDDPKGGQLRAAVYPRLRAH--FQFNELNLFVEVDHHAFTSSNIYSASPST1126
QY	914	----QLSNI-----KESTSFKAKFMIIQSS---DNILKEITRD944
Db	1127	VGFHEISNLVAPQTIDACFEHSGSDIPGLKDIESEGKUKVWNWTSGHRSRLSIATHE1186
QY	945	LK-----DSK-DDAYKG-----IELNINQIKLSPIQESIEEFKDNBEFTL984
Db	1187	LELFARLYDSEGTPAWGQARLPALHAKOLVALEKFANQNRLGDLO-----GOVFST1238
QY	985	INKWFKSFSALGBGYI-----DFKGLDPSIKVRKSLKCKNNKL1026
Db	1239	V--MFEDETYAQRDGTTILROTFFPDQSQQWLSPGFHFVG-TPFYKTPRENTCLNSDYDCL1295
QY	1027	FLYSGANIHQFNRFPEDDKAKSSK---LLMID-KEDLEKVLTKDONVOYOTERVFVRAI1081
Db	1296	DLLTLPDDYLPRWTNYPCADQAQEVAKTPCVTWTELAEDPEPKVT--DY-RLAIRAM1350
QY	1082	ASNTNERTMTISTLSPG-----NCYCNVSIIYNDEKTPISLYKCLKFIISFNSVFDFP1133
Db	1351	LAQSGERTLISAIVPEISHMNVAHSVSYCGSSQNLLLEHSGMCF-----SLPDFE1399
QY	1134	LLRRFVDSNV1143
Db	1400	ICKSTGKANL1409
 RESULT 13 B86128 hypothetical protein Z5900 [imported] - Escherichia coli (strain O157:H7, substrain EDL933) C;Species: Escherichia coli C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004 R;Accession: B86128 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimatauta, E.; Pocamousis, K.; Apodaca, Nature 409, 529-533, 2001 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7. A;Reference number: A85480; MUID:21074935; PMID:11206551 A;Accession: B86128 A>Status: preliminary A:Molecule type: DNA A;Residues: 1-1644 <STO> A;Cross-references: UNIPROT:Q8XC72; UNIPARC:UPI000016SA01; GB:AE005174; MID:g12519308; H: A;Experimental source: strain O157:H7, substrain EDL933 C;Genetics: A;Gene: Z5900		
 Query Match 5.7%; Score 371.5; DB 2; Length 1644; Best Local Similarity 17.7%; Pred. No. 6.7e-07; Matches 246; Conservative 231; Mismatches 470; Indels 443; Gaps 51;		
QY	74	QKAGQIGVESRVVDILLFENDKDKSPNNKLEAKKNNEPIPTIEDIL---LIAEVKRPTF130
Db	143	QEGETLDPLALSLLTAQFPADTDKHKHSURKA--NGEYRSWQDLSTAVTFQNEPPRF200
QY	131	SFDAKDKLKESEDQLYRYLNOYQKHGYLSNGKVWRLYDKSKVLYGEKRYIEFNFXXKIEE190
Db	201	V-----LLGNQLLLDRTK-WAQNRLLRFDFFEELS232
QY	191	KERYEQEWFLFYILIRKERYLKTSNVIWEKEQIAKEKEI--IQKTLKEILIYERPDSS248
Db	233	R--RETDTLKATAVLLHKDSLPGSGAPYLDSDNNSHKHAFGVSDLYALRE-----284
QY	249	IVEFKIAGNYDKPFKSGKITOHIL-----ASTLESIIIFILRIFFIAY293
Db	285	-----SIELLCNEMHYLIDRGLANTYGNRAVPDELSELRYWRLFLFY332
QY	294	IEDNDIFPKILOENKLYRSSISFRY-----FFYDENTKCKLEYKKIITIIFNLLDKGS345
Db	333	IEARPGLGAPMTAKTYLQGSYLETRDLDEMIPLTSEEDNGRGYFHDSLNMLFKLVREG-391

346	Qy	346	DAIKFPVNGG-----LPSDEKVKVYLNNEGLLSISEI-----BILVQMLPFEEK	330
392	Db	392	-----YGGVKWQDLESGBRITITSHQ--FSPVRLESHLPDANNTIRLNVFRNET	442
391	Qy	391	-----NIKDEFVKYSRLDPKPSFGELYETLLLEYDLRIADTTVHRI----	430
443	Db	443	LQOIQAMLSRPAKGRFNRGRISYRQIGINQGVAYEALLSYRGFFASEDIYVKKAG	502
431	Qy	431	-----JEDGVYLIRT-----BEELENKKVNKIATYKGNIVYLTSLRDKKSGAYYTPD	479
503	Db	503	EEFNELETGVYFSKDBISKYHEDEKVYEGDGLSRRIHRKGSFTYRMAGRDRKSASYTPR	562
480	Qy	480	DLTDFWVVISIE-----OLKYKSPDLDKIIDNSCGSHFLISCLDYLTEKWV	527
563	Db	563	VLTRSLVKVALKELFKEQIDPISDPHAKADAILNLTVCBPAMGSAAPLNEAINQLAEAYL	632
528	Qy	528	YELDKPEDVKELDEYRVVIBESBEYVDQDSISKELV-LKRWMLLKCCIYGVNDINPISVE	586
623	Db	623	PH-----KQOAEGRRI-----PQDRVYQELQRVWYIADNNVFGVDLNPVAVE	665
587	Qy	587	ITWLSJWINTFPFGTPLPSFIEHHIKAGNALLG-----YTKDEFF-----	625
666	Db	666	LAEVSJLWLAISGDAFVPFGYHQCGLGSLVAGARRQVFNKSELTYKKAKDPWSLNEPVE	725
626	Qy	626	-----DJKKKFESGFSLEKKRIKEI-----ITIL	650
726	Db	726	LAMNTPREETOIFHFLLPDGGHANYSDKVYKQYPPDFKALDSWRKEFKPSFAGHIADV	785
651	Qy	651	EDIYQIKINGINTTKEDIE-----KSKKIYKEYBESKD	683
786	Db	786	QRISEKVEALWNTYRQOLKAERLKTADSYVWVPAENSEQTRSSLSSKDETFSGRLEDNSA	845
684	Qy	684	IDNRIIFSL-IKLYSLSPDKSLNMEFSDIASVISLIENILGNKTSSEDEKIEK-----	737
846	Db	846	YQKLRWMDYWCALWFWPIDKA--DSLPRDGTWLFIEITLLOGIVITEKYVTEAHTGTD	903
738	Qy	738	-----IRKLSSYY-----KFFHYGIEFPDI---QE	759
904	Db	904	LFAEGLLARESSLFSVAGRLKTEVLFRHLPLAIVDALRKQHRFFHWDLEFCDLFAERG	963
760	Qy	760	GFDIVIGNPPWEKTKNETEFSKHPNY--RKLGIKBEONI KQETLSKDNHPLSIEYNE	817
964	Db	964	GFDMLGNPWLKVVEQWAGVLGDYSPFVLRLKSASLKLATRIDTF-----	1010
818	Qy	818	EKXSITAINNIYKDFK-C-----FTSGGDPNLPFRYVTFNLKLIKKEGNL	862
1011	Db	1011	--NQIPALEAAWSEVEGEGHQMFLNAQNPVPLRGVQTNLYKCFLPQAWRLGAQKGVA	1068
863	Qy	863	TYLVPISAINNESSRILRKHIFARYKLANIYQFENKRF--KDVHSSPKFAJP-----	913
1069	Db	1069	GFLHPEGIYDDPKGGQLRAAVYPRLAH--FQFQNELNLFVEVDHIAKFSNNIYASPSPT	1126
914	Qy	914	-----QLSNI-----KESTTSFKAKFIQSS--DNILKEITRD	944
1127	Db	1127	VGFEHISNLVAPOTIDACFEHSGSGDIPGLKDBIESEGLKVVMWNTSGHRSLSIATHE	1186
945	Qy	945	LK-----DSK-DDAYKG-----IBELNINQIKLSPIOESITIEFKDNBEFTL	984
1187	Db	1187	LELFARLYDSEGTPAQCARLPALHAKQLVAVLEKFAQNPRLGDLQ-----GQVYST	1238
985	Qy	985	INQWFKSFMALGBGYI-----DPFKGLDPSIKWRKSLKKECNKNLI	1026
1239	Db	1239	V--MFDETVQAQRDGTILRQTQFPDSSQWVLSGFHFVG--TPPYKTPRENTCLNSDYDCL	1295
1027	Qy	1027	FLYSGGANIHQFNKRKFEDKDAKESK-----LLMID--KEDLEKVLTKDNOYOTERVFYRAI	1081
1296	Db	1296	DLILTPDDLYLPRNTYIPACDAQBYAKTPCVTWTELAEDPKVTV--DYY--RLAIRAM	1350
1082	Qy	1082	ASNTNERTWISLSPG-----NCYCVNSIYINDEKTPISLYKKLFIISFNFSVFDF	1133
1351	Db	1351	LAQSGRRTISATYPPPEISHMAVRYCYSSQNLLHSGMCF-----SLPDDF	1399

Query Match		5.5%; Score 359; DB 2; Length 2166;
Best Local Similarity		21.4%; Pred. No. 2.8e-06;
Matches		325; Conservative 245; Mismatches 473; Indels 474; Gaps 87;
QY	2 KTNIDVKTNNPNISLYKQSKDPFKENI-----NKLKDPFILNKLFSDIDNST	52
DB	769 RQTEIHK--DKSVFMLEDLNNKFFDDKNFVSKIBECQYKLDKFKIESDILNFKSDLN	826
QY	53 ENIESLLKYIFEELNYSVEQKAGIEGVSERV--DILLFENDKDKASFNNKLK-----	105
DB	827 EF-IESKLQIV---SNIKSDNQK--QIDDFLDRIKSDIL---NRKD--SINNEVDSKLS	875
QY	106 -BAKKNNEPIEDILIIAEVRKPTFSFADAKLKE--SEDOLVRY---LNOYQKHVG	157
DB	876 WQSKLNEITVKLENLSSGVLDLIDSEVTYKIKELFSESLSYYLEKIDEFNQA	935
QY	158 ILSNGKWRVLYDKSVLYGKRYIEFNFXK-----IEEKERY-KQEQWFVLFIYIRKER	211
DB	936 IYSDE---LQDIDMNHFNKETRELEENLSKKFAAVLNNSEEFVKEVD---SLQDKR	986
QY	212 -----YLKTSNV--TEVEKEQIAKEKEII-----QKTLKEIILYE---	243
DB	987 TDIASFQANIDITLDSLVNKFNDINKEINGKYNEVISNYRGSSENISSKLENEIMHEIEN	1046
QY	244 ---RPDSDI-----VFKIAKNIYDKEPKLKGKEIT-----	270
DB	1047 LSRRLTDRIDSLSKGMENLQKLKSFVSKYQVEK-FELVKYDLDTDDGEAKINKLVKEI	1105
QY	271 OHILASILESIIFILIRPIFIAYIEDNDI-----FKKILQBNKLYRSSISPRYFFYDEN	324
DB	1106 EQYKSRLEEAIDYRTI-----DNDIMQAKERFGEITNELK--NNIESKSEFLNDL	1155
QY	325 TKKLEYKKIITIFNLLDKGSDAIFPVNGGLFSDKVLYLNNEGLLSISIEBILVKM	384
DB	1156 YKER-----PKLIESNFEE-RYSTF-----LIESEG--AISKIRDEIYKT	1192
QY	385 LPFEEKNIKDEKPVKYSRLDPSFGELYE---TLLEYDLRIAD-----	424
DB	1193 LITSNDENLO---IKISEMD-QNF-EILQORSKDILFEKELODKIKDCYGFINSQFGEI	1246
QY	425 -----TTVHRIEDGVYLIRTEEELENKVNKIATYALKNTYLTSRSLD	468
DB	1247 KAGVERNINKHFDVCIKKVNTLIDDDI--VKYENEI-HKRIDSLKS-IESTFDSIEKNLN	1302
QY	469 RKSGAYTTPDDLTDPMWTSSI-----EEOQKTSPLD--IKJIDNSCGSGHFLISCL	519
DB	1303 DKVSGCI---DKIANDFNLYIELEERCNEGQNLNENKIDNKIKAIDN-----LALSQY	1353
QY	520 DYLTEKVMVYELDKFEDVKKEDEYRVIIIEESEYDVQDSISKELV--LKRMLLXXCIYG	577
DB	1354 DGLERKY---ADMYDEFSERLNSYIATLSEEPK-----SSNKEMIFELESQKLNKNLE	1404
QY	578 VDNIPISVEITMLSLWINTFIPTPLSFTEHHIKAGNALLGYTKDBEFFDIVKVKFESGFS	637
DB	1405 SDLN--NVEKDVIRLKEEY-----HNVSSHLLKLL---EEDFFKDLKIRGEE---	1446
QY	638 LFKRKRIKEIITILEDIYQKINGINTTKEDIKSKKIYKEYESKDIDNLRIFSLIKL-	696
DB	1447 -LKYSLENFIASYNDKIQNL-----EYDLSKNLENKTELIOQSFRLD	1486
QY	697 -----YLSFPDKSLNMEFSDIASVISIENILGNKTSSEDEKIEKIRKLSSYY	745
DB	1487 IQQKMKDDKENFYLDTKBFSSKKQDMQSEIALMET-----NITGKVD-EFVDFVNNKQSI	1542
QY	746 KFFHYGIEFPDIOEGPDIIVIGNPPWEKTKFNTEFPFSKHIPNRYKLGIRK--EQNI	803
DB	1543 DSWFLNIK--DDVKD-----WQEKSYSTIE---KEI-NLAEIGIKSFENDIFNVKI	1587
QY	804 ----LSKDNHPLSIE--YNEEKSIITAINNIYKDFPKCFTSGGDPNL-----FRYFTFN	854
DB	1588 GLESPDKGFEIKAEEIFSNLQNEAKKIEQSVHLDKF--NIGESLNLKVLDEKFDVDFKLE	1645
QY	855 LIKEKGNL---TYLVPSAIWNNESSRILRKHFARYKLVNIYQFE-----NKKRF	901

DB	1646 KIDEKVNKKTEDILIIQASVKFLTQOKOLEDKIF---ELNQKLEHEFTTSSNLDKVRREM	1702
QY	902 KDVHSSFKFAI-FQLSNIKESTSSFKAKFM-----IQSS-----DNILKEITRDLKSKD	950
DB	1703 VDVISSDKESFEGQTEILINKISBFSEKISLYRNNIETSIENEVNSPFSKISKDLGLLED	1762
QY	951 DAYGIELNINQIKKL-SPIQESI-----IEFK-----DNEEFTLINKM-----F	989
DB	1763 ELKXSLRHSSTSEIETIKSGLOEQDKDFEVEFKONHKELLKEVDNNILESKILMCDVOF	1822
QY	990 SKF-SALGEGYIDFKKGL-----DPSIKNRKSLKKECNKNKILFLYSGAN	1033
DB	1823 NKFISEIKDNLVEYKSDLRAEFEDSYDKINFQIENQIENFKKLDELEKNNSIFL-----	1877
QY	1034 IHQFNRSFFEDKDAKESKLLWIDKEDLEKVL-----TKDNOYQTERVYFVYRAIASNTW	1088
DB	1878 -----EAYSCLKDKLEKLW--ETLKNBIEGLAQEYKNNFENVNKEFY-----NIQKE	1920
QY	1089 TMISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISIENSP--VPDF-----LLR	1136
DB	1921 TL-----GIIEIFNELKLEQES-----IKSIKDNFNRFFEFYSSFSDRYKSLIB	1964
QY	1137 RFVDSNVLKSCLYOCMPQPEEKEILSNPLYLNLAKNTSLLIAKNDPENFKYLLYLEYFK	1196
DB	1965 SYDEMQUIYKAKIKEI---ADEQRTILDN--YERISNKESIL-----	2000
QY	1197 FDKKVNKILKLDKEDFEFKENENNFIASLYSLAKEDFITLLGDPKALKNKKKGBDY	1256
DB	2001 -----KSTIESVDKNFDLINEVEKRFN-----NLSKE-----SAKIQDNLKMENV	2041
QY	1257 ISSLI--KGYNLYLLNN	1271
DB	2042 VSSLLLNKGLSEEVLIN	2058

Search completed: January 24, 2006, 19:59:13
Job time : 55.4458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 274.4 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-16
Perfect score: 6538
Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIKGYDNYLLNKNKIFVHK 1277

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	6534	99.9	1277	2	O50698 BORBU
2	5599	85.6	1278	2	O50667 BORBU
3	5472	83.7	1194	2	O5XYT2 BORGA
4	3119	47.7	806	2	O5XVG5 BORGA
5	2648	40.5	635	2	O5XYI7 BORGA
6	1781	27.0	1250	2	O5H911 CAMJR
7	1764.5	27.0	1250	2	O5H911 CAMJR
8	1648.5	25.2	422	2	O50806 BORBU
9	671	10.3	149	2	O5XYI6 BORGA
10	605.5	9.3	1318	2	O4HA54_9DEIO
11	604	9.2	1442	2	O7NGZ7 GLOVI
12	558.5	8.5	706	2	O52000 HALSA
13	546.5	8.4	662	2	O9HI06 HALSA
14	546.5	8.4	1610	2	O5F0U8 GLUOX
15	526.5	8.1	1181	1	Y3402_METJA
16	525.5	8.0	1186	2	O4HI25 CAMLA
17	522.5	8.0	1257	2	O5HXK7 CAMJR
18	510.5	7.8	1255	2	O4HJ16 CAMLA
19	488	7.5	878	2	O4HEH3 CAMCO
20	473	7.2	134	2	O50807 BORBU
21	455.5	7.0	1252	2	O9ZJ44_HELPJ
22	446	6.8	867	2	O9PU80 CAMJE
23	445	6.8	1203	2	O8E457 STRA3
24	435	6.7	1532	2	O8NSD8 CORGL
25	433	6.6	1532	2	O6M742 CORGL
26	419.5	6.4	1279	2	O26046_HELPY
27	405.5	6.2	5729	2	O8I622 PLAF7
28	398	6.1	2664	2	O7RELO PLAYO
29	395.5	6.0	2723	2	O7RQB6 PLAYO
30	395	6.0	2965	2	O8WRS5 PLAF6
31	394	6.0	2771	2	O26216_9APIC

32	392	6.0	2957	2	O8WRS6 PLAF6
33	392	6.0	5779	2	O8IBS0 PLAF7
34	391	6.0	2770	2	O7YUE9 PLAF6
35	390	6.0	2976	2	O8WRS4 PLAF6
36	389.5	6.0	1513	2	O4YR36 PLAF6
37	388	5.9	2969	2	O7KF73 PLAF6
38	386.5	5.9	2977	2	O8WEP9 PLAF6
39	385.5	5.9	2749	2	O7REY3 PLAF6
40	385.5	5.9	2752	2	O9BJY0 PLAYO
41	382	5.8	1009	2	O8KRW6 BACPU
42	379.5	5.8	1414	2	O63WX7 BURPS
43	379	5.8	2881	2	O6YA78 PLAF6
44	378	5.8	1524	2	O8FSZ4 COREF
45	378	5.8	1993	2	O8ILC6 PLAF7

ALIGNMENTS

RESULT 1
O50698 BORBU PRELIMINARY; PRT; 1277 AA.
ID O50698 BORBU PRELIMINARY; PRT; 1277 AA.
AC O50698;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein BBE02.
GN OrderedLocusNames=BBE02;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp25.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_taxid=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang M.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
"Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."
RL Nature 390:580-586 (1997).
DR EMBL; AE000785; AAC66031.1; -; Genomic_DNA.
DR PIR; E70224; E70224.
DR HSSP; Q57997; 1MJH.
DR TIGR; BBE02; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; F:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6 mtfase.
DR InterPro; IPR002052; N6 Mtfase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PRO0507; N12N6MTFRASE.
DR PROSITE; PS00092; N6 MTASE; UNKNOWN 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 1277 AA; 150972 MW; 0B41DF4EDB5859F4 CRC64;

Query Match 99.9%; Score 6534; DB 2; Length 1277;
Best Local Similarity 100.0%; Pred. No. 1.5e-243;
Matches 1277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKTNDIVKTNPNISLYKQSKDFIKENINKKDFILIKKLFSDIDNSTEAMIESLL 60
DB 1 MKTNDIVKTNPNISLYKQSKDFIKENINKKDFILIKKLFSDIDNSTEAMIESLL 60
QY 61 KYIFELNYSVQKAGQTEGVSRVDILLFENDKDKASFNNKKEAKQNPPIEDIL 120

```
Db 61 KTIIFELNYSVEQKAGQIEGVESRDVILLFENDKDKASFNNKLKEAKQNNPEIPEDIL 120
Qy 121 IIAEVKRPFTSDAKDKLSESDQLYRYLNOYQKHGILSNKGVWRLYDKSVLYGKRY 180
Db 121 IIAEVKRPFTSDAKDKLSESDQLYRYLNOYQKHGILSNKGVWRLYDKSVLYGKRY 180
Qy 181 IEFNFXXKIEBEYEQEWFVFIYLIRKERYLKTNSVIEVEKEQIAKEKIIQKTLKEI 240
Db 181 IEFNFXXKIEBEYEQEWFVFIYLIRKERYLKTNSVIEVEKEQIAKEKIIQKTLKEI 240
Qy 241 LYERPDSDIVFKIANNIYDKFKLSGKEITQHILASILEESIIFILRPIFFIAYIEDNDIP 300
Db 241 LYERPDSDIVFKIANNIYDKFKLSGKEITQHILASILEESIIFILRPIFFIAYIEDNDIP 300
Qy 301 KKLQENKLYRSSISPRFFYDENTKKLEYKKIITIFNLLDKGSDAKFPFVNGGLPSE 360
Db 301 KKLQENKLYRSSISPRFFYDENTKKLEYKKIITIFNLLDKGSDAKFPFVNGGLPSE 360
Qy 361 DKVKYLNNEGLLSISEIEILVKMLFFEEKNKIDKFKVKYSLRDPKPSFGELYETLLEVDL 420
Db 361 DKVKYLNNEGLLSISEIEILVKMLFFEEKNKIDKFKVKYSLRDPKPSFGELYETLLEVDL 420
Qy 421 RIADTTVHRIIEDGVYVIRTEBELENKVKNIATYLGKNIYLTSLDRKKS GAYYTPDD 480
Db 421 RIADTTVHRIIEDGVYVIRTEBELENKVKNIATYLGKNIYLTSLDRKKS GAYYTPDD 480
Qy 481 LTFDFWISSIEBQTKSPDLIKIDNSCGSGHFLISCLDYLTETKVYELDFEDVKKEL 540
Db 481 LTFDFWISSIEBQTKSPDLIKIDNSCGSGHFLISCLDYLTETKVYELDFEDVKKEL 540
Qy 541 DREYRVIISESEYDVQDSISKELVLRMLLKKXCYGVNDINPISVEITWLSWINTFI 600
Db 541 DREYRVIISESEYDVQDSISKELVLRMLLKKXCYGVNDINPISVEITWLSWINTFI 600
Qy 601 TPLSFIEHHIKAGNALLGYTKDEFFDVIVKXKPSGSLFKPKRIKIIITILEDIYQKIGI 660
Db 601 TPLSFIEHHIKAGNALLGYTKDEFFDVIVKXKPSGSLFKPKRIKIIITILEDIYQKIGI 660
Qy 661 NDTTKEDIEKSKIKIYKVEESKDIDNRIIFSLIKLISLSPDKSLNMFESDIASVISLIE 720
Db 661 NDTTKEDIEKSKIKIYKVEESKDIDNRIIFSLIKLISLSPDKSLNMFESDIASVISLIE 720
Qy 721 NILGNKTSSEDEKIEKIRKLSYKFFHYGIEFFDIOEGFDIVIGNPWEKTKNETEF 780
Db 721 NILGNKTSSEDEKIEKIRKLSYKFFHYGIEFFDIOEGFDIVIGNPWEKTKNETEF 780
Qy 781 FSKHIPNVRKLGIEQNIITKQIILSKDNHPLSIEYNEEKNSIIANNIYKFDKFTSGG 840
Db 781 FSKHIPNVRKLGIEQNIITKQIILSKDNHPLSIEYNEEKNSIIANNIYKFDKFTSGG 840
Qy 841 DPNLFRYFTFNLKIKKEGNLTLYLPSAIWNSSRLRKHI FARYKLYNYIQFENKRR 900
Db 841 DPNLFRYFTFNLKIKKEGNLTLYLPSAIWNSSRLRKHI FARYKLYNYIQFENKRR 900
Qy 901 FKDVHSSFKFAIFQLSNIKESTSSFKAKPMIOSSNLIKETRLDKSDKDAYKGIELNI 960
Db 901 FKDVHSSFKFAIFQLSNIKESTSSFKAKPMIOSSNLIKETRLDKSDKDAYKGIELNI 960
Qy 961 NQIKKLSPIQESIIEBFKONEEFTLINMPSKFSALGEGYIDFKKGLDPSIKNRKSLKREC 1020
Db 961 NQIKKLSPIQESIIEBFKONEEFTLINMPSKFSALGEGYIDFKKGLDPSIKNRKSLKREC 1020
Qy 1021 NNKNLIFLYSGANIHFQNSRPFEDKDAKESKLLWIDKEDLEKVLTKONQYTERVFYRA 1080
Db 1021 NNKNLIFLYSGANIHFQNSRPFEDKDAKESKLLWIDKEDLEKVLTKONQYTERVFYRA 1080
Qy 1081 IASNTNERMTISTLSPGNCYCNSIYINDEKTPISLYKKLFIISFNPSVDFLLRRFVD 1140
Db 1081 IASNTNERMTISTLSPGNCYCNSIYINDEKTPISLYKKLFIISFNPSVDFLLRRFVD 1140
Qy 1141 SNVLKSCLYQCPMPQPEBEKILSNPLYLNIAKNWTSLLIAKNDPENFKYLLYLEYKFDKE 1200
```

```
Db 1141 SNVLKSCLYQCPMPQPEBEKILSNPLYLNIAKNWTSLLIAKNDPENFKYLLYLEYKFDKE 1200
Qy 1201 KVNKILKLDKDEDFEKKKENENNFIISLASYSLAKEDFTITLLGDFKALKNKKGGEDYISL 1260
Db 1201 KVNKILKLDKDEDFEKKKENENNFIISLASYSLAKEDFTITLLGDFKALKNKKGGEDYISL 1260
Qy 1261 IKGYDNYLLNNKFIYHK 1277
Db 1261 IKGYDNYLLNNKFIYHK 1277

RESULT 2
OS0667 BORBU PRELIMINARY; PRT; 1278 AA.
AC OS0667
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Hypothetical protein BBH09.
GN OrderedLocusNames=BBH09;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteriia; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI TaxID=1139;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Cagjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-P., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Wattley L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi";
RL Nature 390:580-586 (1997).
DR EMBL; AB000784; AAC66000.1; -; Genomic_DNA.
DR PIR; B70236; B70236.
DR TIGR; BBH09; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 1278 AA; 150871 MW; 8A0B22DF16655C0 CRC64;

Query Match 85.6%; Score 5599; DB 2; Length 1278;
Best Local Similarity 85.3%; Pred. No. 1.5e-207;
Matches 1082; Conservative 94; Mismatches 89; Indels 4; Gaps 3;

Qy 6 IVKTNPNISLYKQSKDFIKKENINKLKFILIKNKLFSIDDNSTANIESLKYIPE 65
Db 12 IIKTNPNVSLYKELSGKFIKKENIVKSKNFFIFLKNKIQAIDDNSTANIESLKSIFE 71
Qy 66 ELNYSVEQKAGQIEGVESRDVILLFENDKDKASFNNKLKEAKQNNPEIPEDILIIAEV 125
Db 72 ELAYSVEQKGGQIEGVKSRVDILLFENDKDKVAFNKKLEAKQNNPEIPTEDILIIAEV 131
Qy 126 KRPTFSFPAKDKLSESDQLYRYLNOYQKHGILSNKGVWRLYDKSVLYGKRYIEFN 185
Db 132 KRPSFSPDAKDKVAEADQLYRYLNOYQKHGILSNKGVWRLYDKSVLYGKRYIEFDF 191
Qy 186 XKIEBEYKQEWPFVFIYLIRKERYLKTNSVIEVEKEQIAKKEIIOKTLKELYRPP 245
```

Db 192 SKIKEBYKEQWFIILSYLIRKERYLTSNIIISVEKEQISKEKEIIQKTLREILYERP 251
Qy 246 DDSIVFKIAKNIYDKFKLQSGKEITQHILASILEESIIFILRIFFIAYIEDNDIPFKILQ 305
Db 252 DDSIVFKIAKNIYDKFKLQSGKEITQHILASILEESIIFILRIFFIAYIEDNDIPFKILE 311
Qy 306 ENKLYRSSISPRYPFYDENTKKLEYKIIITFNLLDKGSDAIKPPVNGGLPSEDVKY 365
Db 312 ENKLYRSSISPRYPFYDENTKKLEYKIIITFNLLDKGSDAIKPPVNGGLPSEDVKY 371
Qy 366 LNNEGLLSISEEILVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDLRIADT 425
Db 372 LNNEGLLSISEEILVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDLRIADT 431
Qy 426 TVHRIEDGVYLIRTEBELENKKNKIATYKGNIIYLSRSLDKKSGAYTTPDPLTDFM 485
Db 432 TVHRIEDGVYLIRTEBELENKKNKIATYKGNIIYLSRSLDKKSGAYTTPDPLTDFM 491
Qy 486 VISSTEEOLKTKSPDLIKIIONSCGSHPLISCLDYLTEKVMYELDKEDVKELDEBYR 545
Db 492 VISSTEEOLKTKSPDLIKIIONSCGSHPLISCLDYLTEKVMYELDKEDVKELDEBYG 551
Qy 546 VIESESYVDQDSISKELVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDLRIADT 605
Db 552 VIESESYVDQDSISKELVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDLRIADT 611
Qy 606 IEHHIKAGNALLGYTKDBFFDIIVKKGPSGSLFKPKRIKEIITILEDIYQKINGINDTTK 665
Db 612 IEHHIKAGNALLGYTKDBFFDIIVKKGPSGSLFKPKRIKEIITILEDIYQKINGINDTTK 671
Qy 666 EDIEKSKIIYKYEESKIDNRIIFSLIKLISLSPDKSLNMFSDIASVLSLNIENLGN 725
Db 672 EDIEKSKIIYKYEESKIDNRIIFSLIKLISLSPDKSLNMFSDIASVLSLNIENLGN 731
Qy 726 KTSSEDEKIEKIRKLSYKFFHYHIEFPDIOEGFDIVIGNPPWEKTFNETEFPFSKHI 785
Db 732 KTSSEDEKIEKIRKLSYKFFHYHIEFPDIOEGFDIVIGNPPWEKTFNETEFPFSKHI 791
Qy 786 PNYRKLGIKEQNIIOEILSKDNHPLSIEYNEEKNSSIIAINNIYKDFKCTSGGDPNLF 845
Db 792 PNYRKLGIKEQNIIOEILSKDNHPLSIEYNEEKNSSIIAINNIYKDFKCTSGGDPNLF 851
Qy 846 RYFVTFNKLKEKGNLTVLPSALWNESSRILRKHIFARYKLNIIYIOPENKKGPKDVH 905
Db 852 RYFVTFNKLKEKGNLTVLPSALWNESSRILRKHIFARYKLNIIYIOPENKKGPKDVH 911
Qy 906 SSFKAIFQLSNIKESTSEKAKFMIQSDNLIKETIOLKDSKDDAYKGIELNINOLKK 965
Db 912 SSFKAIFQLSNIKESTSEKAKFMIQSDNLIKETIOLKDSKDDAYKGIELNINOLKK 971
Qy 966 LSPQIESIIEFKDNEEFTLINKMFSKFGALGEGYIDFKKGLDPSYKRNKSLKKECNKNL 1025
Db 972 LSPQIESIIEFKDNEEFTLINKMFSKFGALGEGYIDFKKGLDPSYKRNKSLKKECNKNL 1029
Qy 1026 IFLYGANIHQNSRFFEDKAKESKULWIDKEDKVLTKDNQYQTERVYFRAISNT 1085
Db 1030 IFLYGANIHQNSRFFEDKAKESKULWIDKEDKVLTKDNQYQTERVYFRAISNT 1089
Qy 1086 NERTWISTLSPGNCYCVNSIYINDEKTPISLYKLPISISFNSFVDFELLRRF-VDSNVL 1144
Db 1090 NERTWISTLSPGNCYCVNSIYINDEKTPISLYKLPISISFNSFVDFELLRRF-VDSNVL 1149
Qy 1145 KSCLQCPMPQPEEKEILSNPLYLAKNTSLIIAKNDPENPKYLLLYEYFKFDEKVNK 1204
Db 1150 KSCLQCPMPQPEEKEILSNPLYLAKNTSLIIAKNDPENPKYLLLYEYFKFDEKVNK 1209
Qy 1205 ILKLDKEDBFFKEKENENNFIITASIYSLAKEDFITLLGDFKALKNKKGEDYISSLIKGY 1264
Db 1210 ILKLDKEDBFFKEKENENNFIITASIYSLAKEDFITLLGDFKALKNKKGEDYISSLIKGY 1268
Qy 1265 DNYLANKI 1273
Db 1269 ENYLKTNKL 1277

RESULT 3

Q5XYT2 BORGA PRELIMINARY; PRT; 1194 AA.
AC Q5XYT2; 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames=BGP199;
OS Borrelia garinii PBI.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_taxid=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Schilhabel M., Lehmann R., Platzer M.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY722925; AAU86050.1; -; Genomic_DNA.
GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; P:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6 mtfase.
DR InterPro; IPR002052; N6 Mtfase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PR00507; N12N6MTFASR.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 1194 AA; 141642 MW; 51DC343D205F2B41 CRC64;
Query Match 83.7%; Score 5472; DB 2; Length 1194;
Best Local Similarity 88.5%; Pred. No. 1.1e-202;
Matches 1056; Conservative 74; Mismatches 63; Indels 0; Gaps 0;
Qy 1 MKTNDIVKTNPNISLYKQLSKDFIKENINKLKDFFILIKNKLFSIDNTEANIESLL 60
Db 1 MKTNNIVKTNPNISLYKQLSKDFIKENINKLKDFFILIKNKLSSIDNTEANIESLL 60
Qy 61 KYIFEELNYSVEQKAGIEGVESRVDILLFENDKDKASFNKLKEAKNNEPIEDIL 120
Db 61 RSIFEELNYSVEQKAGIEGVESRVDILLFENDKDYRVDFFNNKLEAKNNEPIEDIL 120
Qy 121 IIAEVKRPFSFADAKDLKESDQLYRYLNOYQHYGILSNQKWLRYDKSKVLYGEKRY 180
Db 121 IIAEVKRPFSFADAKDLKESDQLYRYLNOYQHYGILSNQKWLRYDKSKVLYGEKRY 180
Qy 121 IIAEVKRPFSFADAKDLKESDQLYRYLNOYQHYGILSNQKWLRYDKSKVLYGEKRY 180
Db 121 IIAEVKRPFSFADAKDLKESDQLYRYLNOYQHYGILSNQKWLRYDKSKVLYGEKRY 180
Qy 181 IEFNFKXIEEKEEYKQEWFLVILIRKERYLKTNSVIEKEQIAKEKEIIQKTLREI 240
Db 181 IEFNFKXIEEKEEYKQEWFLVILIRKERYLKTNSVIEKEQIAKEKEIIQKTLREI 240
Qy 241 LYERPDDSDIVFKIAKNIYDKFKLQSGKEITQHILASILEESIIFILRIFFIAYIEDNDIF 300
Db 241 LYERPDDSDIVFKIAKNIYDKFKLQSGKEITQHILASILEESIIFILRIFFIAYIEDNDIF 300
Qy 301 KKLIOENKLYKSSISFRYFFYDENTKKLEYKIIITFNLLDKGSDAIKPPVNGGLPSE 360
Db 301 KKLIOENKLYKSSISFRYFFYDENTKKLEYKIIITFNLLDKGSDAIKPPVNGGLPSE 360
Qy 361 DKVKYLNNEGLLSISEEILVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDL 420
Db 361 DKVKYLNNEGLLSISEEILVLMPLFEEKNIKDEKFKVYSRDLPKSPFGLYETLLLEYDL 420
Qy 421 RADTTVHRIEDGVYLIRTEBELENKKNKIATYKGNIIYLSRSLDKKSGAYTTPDD 480

Db	421	RIADTTVHRIVBEVYLIRTEEBELNKKVNVATYYKGNITLYLTSRSLNRKKSQAYTPDD	480
Qy	481	LTPDMWISSIEEOLKTKSPLDIKIINDSCSGHFLISCLDYLTEKWYELDKPEDVKEL	540
Db	481	LTPDMWISSIEEOLKTKAPJDIKIINDSCSGHFLISCLDYLTEKWYQDKPEDVKEL	540
Qy	541	DEEYRVIIIESEEYDVODSISKELVLRMLLKCIYGVNDINPISVEITMLSLWINTPIFG	600
Db	541	DKEYRALIKESSEYDVODSISKELVLRMLLKCIYGVNDINPISVEITMLSLWINTPIFG	600
Qy	601	TPLSFIEHHIKAGNALLGYTKDBFFDIVKKKFSGGSFLFKRKIKEIITILEDIYQKIKGI	660
Db	601	TPLSFIEHHIKVGNALLGYTKDBFFDITKKKFGGGSFLFKRKIKEIITILEDIYQKIKGI	660
Qy	661	NDTTKEDIKSKKIYKEYEESKDIDNLRITFLSLIKLYLSLSPDKSLNMFSDIASVLSLIE	720
Db	661	NDTTKEDIKSKKIYKEYEESDNDNLIIFSLIKLYLSLSPDKSLNIEFSDIATVLSLIE	720
Qy	721	NIILGNKTSSDKKEIKIRKLSYKYPFHGIERPDDIOEGFDIVIGNPPMEKTKFNETEP	780
Db	721	NIILNKTTSSDKEIEIRKLSYKYPFHGIERPDDIOEGFDIVIGNPPMEKIKFNESEF	780
Qy	781	FSKHIPNYRKILGIEQNIIEKQELISKDNHPLSEYNEEKNSIITAINNIYFDFKCFCTSGG	840
Db	781	FSKHIPSYRKLSIKEQNKIKEELISKDTHPLNIEYNEEKNSISTINNIYFDFKDYTSGG	840
Qy	841	DPNLFRYFVTNKLIIKEKGNLTYLVPISAIWNSSSSKILRKHIFARYKLYNIYQFENKKR	900
Db	841	DPNLFRYFVTNKLIIKEKGNLTYLVPISALPSSSSSSRLKHHIFSFKLYNIYQFENKKR	900
Qy	901	FKDVHSSFKFAIPQLSNIKESTSPKAKPMIOSSDNLILKEITRDLKDSKDDAYGIELNI	960
Db	901	FKDVDSRPKFVIFQFSNIKESTSPKAKPMIOHSDNLILKEITRDLKDNKODPYGIELNI	960
Qy	961	NOIKKLSPIOESIIEFKDNEEFTLINKWFSEFSGALGEGYIDFKGLDPSIKNRKSLKKEC	1020
Db	961	DQIKUGSPTQESIIEFKNNEBFGLINQWFQSVFSBEYIDFKGLDSTSINKNRKSLKEY	1020
Qy	1021	NNKMLIFLYSGANIHQFNRSFFEDKDAKSSKLIIWDKEDLVLTKDNOYQOTERVFYRA	1080
Db	1021	DNENFIPLYCGANIHQFNRSFFENKNKAKESKFPWIDKEDLEKILAENNOHOTERIFYRA	1080
Qy	1081	IAGNTNERTWISTLSPGNCYCVANSIYINDEKTPISLYKCLFIISFNFSVFDELLRRFVD	1140
Db	1081	IARNTDIRTWISTLSPQNCYCVSMYINYEKTPISIVYKCLFIISILNSLAFDFLLRRFID	1140
Qy	1141	SNVLKSLCYOCMPDQPEKEILSNPLIYLNIAKNTSLLIAKNDPENFKILLYLE	1193
Db	1141	SNVLKSLCYOCMPDQPEKEILSNSLYLNIAKNTSILIAKNDPENFKILLYLD	1193

[illegible]

```
Db 781 YOCPMQPEERKNILANPLYLTLVKNT 806
||||| ||||| ||:||||| |||
Q5XY17_BORGA PRELIMINARY; PRT; 635 AA.
AC Q5XY17_
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames=SGP294;
OS Borrelia garinii PBI.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Glockner G., Lehmann R., Romaldi A., Pradella S.,
RA Schulte-Spechtel U., Schilhaber M., Wileke B., Suhnel J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Nucleic Acids Res. 32:6038-6046(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Schilhaber M., Lehmann R., Platzer M.;
RA Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY722940; AAU86145.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Hypothetical protein.
SQ
SEQUENCE 635 AA; 74562 MW; 0CB490AB4AD4BE18 CRC64;
Query Match 40.5%; Score 2648; DB 2; Length 635;
Best Local Similarity 81.0%; Pred. No. 2.8e-94;
Matches 516; Conservative 44; Mismatches 75; Indels 2; Gaps 1;
QY 569 MLLKXCIYGVNDINPISVEITMLSWINTFIFGTPLSFIEHHIKAGNALLGYTKDEFDIV 628
Db 1 MLLKXCIYGVNDINPISVEITMLSWINTFIFGTPLSFIEHHIKVGNALLGYTKDEFDIT 60
QY 629 KKKFESGSLFKRIKEITILEDIYQIKGINDTTKEDIEKSKIKYKEYESKDIDNLR 688
Db 61 KKKFESGSLFKRIKEITILEDIYQIKSINDTTKEDIEQSKIKYKEYEBENEDIDYLR 120
QY 689 IIFSLIKLYSLSPDKSIAEESDIASVLSLTENILGNKTSSEDEKIEKIRKLSYYKFF 748
Db 121 IIFSLIKLYSLSPDKSIAEESDIASVLSLTENILGNKTSSEDEKIEKIRKLSYYKFF 180
QY 749 HYGIEFPDIQEGFDIVGNPPWEKTKFNETEPFSKHIPNYRKLGIKEQNLIKQELISKDN 808
Db 181 HYGIEFSDIQEGFDIVGNPPWEKTKFNEAEFFSKHISYRKLSEKQNKIEELISKGN 240
QY 809 HPLSIYNEEKNSIATINNIYKDFKFTSGGDPNLFYVTFNLKIKEGNLTYLVP 868
Db 241 HPLSIYNEEKNSISTINNIYKDFKFTSGGDPNLFYVTFNLKIKEGNLTYLPI 300
QY 869 AIWNESSESLRKHPARYKLYNYIYQENKPKDVKHSPFAIPQLSNIKESTSFPKAK 928
Db 301 ALWSESSSRALKHKLINLYNYIYQENKPKDVKHSPFAIPQLSNIKESTSFPKAK 360
QY 929 FMIQSSDNILKEITRDLKSDKDAYKGIENLNQIKLSPIQESIIEPKDNIEFTLIN 988
Db 361 FMIQGNIDNIEKMAWNLKDKGNKAYKGIENLDQIKLSPIQESIIEFPKNKEFTLIN 420
QY 989 FSKFSALGEGYIDPKGLDPSIKNRKSLKKECNKNLIFYSGANIHFNSRFPEDKDAK 1048
```

```
Db 421 FIQFSILSEYIDFIFGL--NLTKNALLKESNNKNFIFYCGANIHFNSRFPFKSVK 478
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1049 ESSKLLMIDKEDLEKVLTKNQYQYTERVYFRAIASNTNERTMISTLSPGNCYCVNSIYN 1108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 479 ESSKSFWDKEDLAKVLTEDNKLNESIIYRKARNTDITRTMISTLCPKNSYCIESLCIN 538
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1109 DEKTPISLYKGLFIISIFNSFVDFLLARRFVDSNVLSCLYQCPMPQPEEKEILSNPLYL 1168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 539 YEKTAISIYKGLFIISIFNSLAFDFLLRRFVDSNVLSCLYQCPMPQPEEKDILANSYLL 598
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1169 NLAKNTSLIAKNDPENFKYLLLYEYFKDFCKEYKNI 1205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 599 TLVNTSLIAKNDPLNFSNLLYLHHPKFSKEKVDKI 635
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
Q9PPL7_CAMJ7 PRELIMINARY; PRT; 1250 AA.
AC Q9PPL7_
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Possible restriction /modification enzyme.
GN OrderedLocusNames=Cj0690c;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10689204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Kettle J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB2964.1; -; Genomic_DNA.
DR FIC; B13339; E81339.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Complete proteome.
SQ
SEQUENCE 1250 AA; 147698 MW; 0A2CD9C5A464CDEA CRC64;
Query Match 27.2%; Score 1781; DB 2; Length 1250;
Best Local Similarity 34.7%; Pred. No. 1.4e-60;
Matches 457; Conservative 244; Mismatches 479; Indels 138; Gaps 40;
QY 5 DIVTNNPNISLYKQLSKDPIKKENIKLKDPFF-----ILIKNKLFSIDDDNSTEAN 55
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6 ELKYLNNLTSTSYSLSDIDFPKIYFNANQAFRDLTKITKIYDNKFKIQNEHQFE-- 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 56 IESLLKYIFEELNYSVEQKAGQIEGVSRSVDILLFENDKDKASFNKKLEAKKNNEP 115
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db -64 -DEFISKVLEILGWCVRQDEKILQGLEKEDFLLPSNDKLKSKYENLDKTKSS---- 118
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 116 IEDILITAEVKRPTFSFDADKLKESD-QLYRYLNQYQKHYGILNSGKWLVDKSKVL 174
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 119 -NDFIILLESKAYNIEIDNK-KVKNDPHFQILRYLGNLKNQYGLTNGRFRFYDNS-IL 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 YGEKRYIEFNPXKIEEKEEYKEQEWFLFYILIRKRYLKTNSVIEVEKEQIAKEK 234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 176 NSNKVFYEINLEKIIEDQNI---EAFAYFYSVFSAFNFTFKEDHLEITLQNNKLSKIKIE 232
QY 235 KTLKEILYERPDSDSVFK-IAKNIYDKFKLSGKEITQHILASILEESIIFILRIFFIAY 293
Db 233 DDLKSIIVGTNGNESLFFIGSRIYNK-----TKADLKLIYENSILYIFIRLLFIAY 283
QY 294 IEDNDIFKKILOENKLYRSSISFRYFF--YDENTKKKLEYKKIIFLNLKDGSDAIKPP 351
Db 284 FEDK--FEILLEKHYPKSKISLRTLLENLQDESSGGFGELENIFFIYNGKGNFDMF 341
QY 352 VFNGLFSDKVKYVNNGLLSISEIEILVKMLPFEKNIKDEKFKYKSLRDPKSGFEL 411
Db 342 VFNGLFSDKSTALLSTPKIFNDKDLKFLINQLNFKDKNLSFKR--DYKTLISVEHLGTI 399
QY 412 YETLLEYDLRIADTTVHRIIDGVVLIIRTEBELE-----NKKVKNKIATYL 456
Db 400 YEGLLSYFPEIAN-----EDIYVSVYKESKEIECVFNDVDFKILEKSKVKETPYK 452
QY 457 KGNIVLTSRLDRKKSAGVYTPDDLTDFMVISSEIEQLKTKSPLDIKIINDSCSGHFLI 516
Db 453 KQVILKNTSNRKSASAPYTPQSTANFLIQSALKDKLNNENILFKILDNACSGHFLV 512
QY 517 SCLDYLTEKVVYELDKFEDVKELDEEYRVIIIESEYDQD-SISKELVLKMLKOCI 575
Db 513 GVLNAITHIVLSDFDHFTNLKELYBEKENILNHIKDF-VQDYEVDESILKRLLLKRII 571
QY 576 YGVNDINPISVEITMLSLMINTIFCTPLSFIEHHIKAGNALLGYTKDEFFDIVKVKFSG 635
Db 572 YGVNDLNPFSIELTKLSWDSIFTPPLSFIEHHIKOGNALLNSNLSDFKDLIK---QNS 628
QY 636 FSLFKKRIKEIITILEDIYQIKINGINDTTKEDIKSKIKIYKESKIDINLRIIFSILK 695
Db 629 SNLFTNSITQBEILQEVPEKLDNLKOTNEBQIKQSKQIYQN-ETLPKLNKLNLVINYIN 687
QY 696 LYSLSFDKSLNMFSDIASVISLIENLGNKTTSEDEKIEKIRKLSYKFFPHYGIBFP 755
Db 688 --TLHF---VKNBEILQIKALS--QDDIQNLQONEQAKAI--ISKYQKEFNPFNYLELFP 738
QY 756 DIQE-----GPDIVGNPWEKTKNETEFPFSKHIPNRYKLGIKQNIKQBIKSLKDNHP 810
Db 739 EIVENQVFGFOIIGNPNWDKTKGSDSDFPQYKSDYRSLSASKKKEIQDNLAKDY-- 796
QY 811 LSIYNEEKNSIIANNIYKPFKCTSGGDNLPFRYFVTENLKLKEKGNITLYVPSAI 870
Db 797 IKQNTKQKAYINDISEYKKAAYPLNKGSGDGNLFLFVEKNSLSLKQDGNLAYVPSAL 856
QY 871 WNESSRILRKHIFARYKLNIYQFENKCR-FKDVHSSFKFAIFOLSNIKESTSSFKAKF 929
Db 857 MFDGSLILRKEILENKTLEYFYSFENNKAIFIDVHRSYKFAIMLIK-N-QANHTHKIM 915
QY 930 MIQSSD-NILKEITRDLKSDDAYKGIETLNQIKLSPIQESIIEPKDNEEFTLINKM 988
Db 916 MFYKTDINSLK-----NKDEI---LTLNLDKIKGUSPTHALMELKQALBILRKS 964
QY 989 FSKFSAIGYIDFKGLDPSIKNRKSLKECNKNKLIPLYSGANTHQNSRF-----FE 1043
Db 965 YNAFQNLSDYIDFREL-D--MTNDKOLFIEFREGLLPDYEGKMLHQFDANFSQATYFL 1022
QY 1044 DK---DAKESKLLMIDKEDLEKVL-T-KDNQYTE--RVFYRAIASNTNERTMISTLSPG 1097
Db 1023 ERAKFDELRKSLYRAKATGKELNPCLIKYDRBFFRIGRYKIASDNTNERTLIASLLPK 1082
QY 1098 NCYCNSIYINDEK-----TPISLYKK-LFIISIFNSFVDFPDLRRFVDSNVLKS 1146
Db 1083 NCGGADSTYSNPKQYVLKDDVICMDIVPEYRILFVLALFNSLVDFIIRNNVQINVSKS 1142
QY 1147 CLYQCPMPQPEKEILSNPLNLAKNNTSLLIAKNDPENFKVLLVLEVPKPKQVKN--- 1203
Db 1143 YLERIPLPQPSDEEQNNEIYKTLAKNALLQIYDQND-----RHPDELKQFENIKN 1194
QY 1204 -KILKLDKDEDFPKEKENENNFIIASLSYSLAKEDFTLLGDPKALNKKKGEDYISSL 1260
Db 1195 BEIPTKKAYDILRAK---NDLLVKEGLSDEFSYMTISFPKVLNEKQ--SEYITLL 1247

RESULT 7

QSHV91 CAMJR QSHV91 CAMJR PRELIMINARY; PRT; 1250 AA.
AC Q5HV91;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=CJ0E0789;
OS Campylobacter jejuni (strain RM1221).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=195099;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15660156; DOI=10.1371/journal.pbio.0030015;
RA Fouts D.E., Mongodin E.F., Mandrell R.E., Miller W.G., Raeko D.A.,
RA Ravel J., Brinkac L.M., DeBoy R.T., Parker C.T., Daugherty S.C.,
RA Dodson R.J., Durkin A.S., Madupu R., Sullivan S.A., Shetty J.U.,
RA Ayodeji M.A., Sivarubsbeyn A., Schatz M.C., Badger J.H., Fraser C.M.,
RA Nelson K.E.;
RT "Major structural differences and novel potential virulence mechanisms
from the genomes of multiple Campylobacter species.";
RL PLoS Biol. 3:72-85(2005).
DR EMBL; CP000025; AAW34576.1; -; Genomic_DNA.
DR TIGR; CJ0E0789; -;
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR007409; DUF450.
DR InterPro; IPR002296; N12N6_mtfrase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF04313; HSDR_N; 1.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1250 AA; 147549 MW; A367AD74887B235 CRC64;
Query Match 27.0%; Score 1764.5; DB 2; Length 1250;
Best Local Similarity 34.1%; Pred. No. 6.1e-60;
Matches 452; Conservative 250; Mismatches 474; Indels 149; Gaps 41;
QY 5 DIVTNNPNISLYQLSKDFIKKENINKLQDFF-----ILIKNKLFIIDDNSTEAN 55
Db 6 ELKYLNNLFTSYSLIEDFPFKIYNFNANQAFRDLTKITKIYDKNFKIQNEHOFE-- 63
QY 56 IESLLKYIFEELNYSVEQKAGQIEGVESRVDILLFENDKDKASFNNKLKKAQKNEPIP 115
Db 64 -DEFISKVLEILGWCVFVRQDEKIIQGLEKDPDFLLFSNDKLKSKYENLDKETKSS--- 118
QY 116 IEDILITAEVVRPTFSFDAKDKLKESED-QLYRYLYNQYKHGYGLNSGVWRLYDKSKVL 174
Db 119 -NDPTIILSKAYNIEIDNK-KVQDNPHQILRYLGNLKNKYGFLTNGRFRWFRYDNS-IL 175
QY 175 YGEKRYIFENPKYKIEEKEEYEQEWFVLIYLRKERYLKTNSVIEVEKEQIAKEEIIQ 234
Db 176 NSNKVFYEINLEKIIEDQNI---EAFAYFYSVFSAFNFTFKEDHLEITLQNNKLSKIKIE 232
QY 235 KTLKEILYERPDSDSVFK-IAKNIYDKFKLSGKEITQHILASILEESIIFILRIFFIAY 293
Db 233 DDLKSIIVGTNGNESLFFIGSRIYNK-----TKADLKLIYENSILYIFIRLLFIAY 283
QY 294 IEDNDIFKKILOENKLYRSSISFRYFF--YDENTKKKLEYKKIIFLNLKDGSDAIKPP 351
Db 284 FEDK--FEILLEKHYPKSKISLRTLLENLQDESSGGFGELENIFFIYNGKGNFDMF 341
QY 352 VFNGLFSDKVKYVNNGLLSISEIEILVKMLPFEKNIKDEKFKYKSLRDPKSGFEL 411
Db 342 VFNGLFSDKSTALLSTPKIFNDKDLKFLINQLNFKDKNLSFKR--DYKTLISVEHLGTI 399
QY 412 YETLLEYDLRIADTTVHRIIDGVVLIIRTEBELE-----NKKVKNKIATYL 456

```
Db 400 YEGLSVFFPEIAN-----EDIIYVYKESKIECYFDNDYDPKILEKSKKVEKYFYK 452
Qy 457 KGNILYTSRSDRKKSGAYTTPDLDLTFWVISSIEEQKLTSPDLIDKIDNSCGSHFLI 516
Db 453 KGQIYLNKTSNSRKASAFYTPQSIFANFLIQSALKDKNLNNILKFKILDNACSGHFLV 512
Qy 517 SCLDYLTKEVWVYELDKFDVVKELDEEVRVIERSEEDVDOD-SISKELVLRMLKXCI 575
Db 513 GVLNAITHVLSDFDHTNLKELYEKENILNHIKOP-VQDYEVDESILKRLLLKRII 571
Qy 576 YGVDINPISVITMLSLMINTFICTPLSFTEHHIKAGNALLGYTKDFFDVIKKPBG 635
Db 572 YGVDLNPISIELTKLSLWIDSFIFTPPLSFTEHHIKAGNALLNSLNSLDFKLIK---QNS 628
Qy 636 FSLFKRIKIEITILEDIYQIKGINDTTKBEDIEKSKIKYKEYESKDIDLNLRIIFSJK 695
Db 629 SNLFNITSITQFEILQVFEKLDNLKOTNEBQIKSQKIYQN-EITPKLKLNLVLYNIN 687
Qy 696 LYSLSFKSLNMFSDIASVISL---IENILGNKTSSEDEKIEKIRKLUSSVYKFFHYGIE 753
Db 688 --TLHF---VNKEELQILKALSQDDIQNL-----SOMEQAKAVISKYQKEFNFFNYELE 736
Qy 754 FPDIOE-----GFDIVIGNPWEKTKENETBFSSKHIPNRYKLGIKIKEQNIIBQELSKDN 808
Db 737 FPEIVENQVFGFDIIIGNPPWDKTKFSDDDFFPOYKSDYRSLSIAASKKEIQDNLAKDY 796
Qy 809 HPLSTEYNEEKNSITAINNIYKDFPKCTSGDGNLFRYFTVFNILKIKERGNITLYLPS 868
Db 797 --IKONYEKQAYINDLSEYKAYPLNKGSDGNLFLVEKNLSLLQDGNLAYVLPS 854
Qy 869 AIWNSSRIIRKHIFARYKLNIIYOFENKCR-FKDVHSSPKFALFQLSNIEKSTSPKA 927
Db 855 ALMPEDGSLTLRKEILENKTLLEYFYSFENRQAFADVHKSFKALMQVKN-TQANHHTKI 913
Qy 928 KFMIOSSD-NILKEITRDLKSDKDAYKGIELNINQIKLSPIQESIIIEPKDNEEFTLIN 986
Db 914 KTMFYKTDWNPJK-----NKDEI---LTLSLKDIKIKLSPTHLALMELKDKQALELR 962
Qy 987 KMFSSFSALGEGYIDFKKGLDPSIKNRKSLKECNKNKLIPLYSGANTHQSRSF----- 1041
Db 963 KYNFAQNLSPYIDFVNEL---HWINDKDLFIEEFREGLLPLFEGKMIHQDFTNFSQATY 1020
Qy 1042 -----FEDK-DAKESKLLMIDKEDLEKVTLDNQYQTE--RVFYRAIASNTNERTMIS 1092
Db 1021 FLEKAKFDERLKSLSRAKATGKELNPKLIK---YDREFRGLGYRTIASDTNERTLIA 1077
Qy 1093 TILSPGNCVCNSIYINDEK-----TPISLYKK-LFISIFNSFVDFELLRFRVDS 1141
Db 1078 SLLPKNCGCNSVYSNIPKQYIVKDDAICMDIVPYERILFVLALFNSLWVDFIIRNMVQI 1137
Qy 1142 NVLKSCLYQCPMPQPEEKEILSNPLYLNLAKNTSLIAKNDPENPKYLLYLEYKFPDKEK 1201
Db 1138 NVSKTYLERIPQPSDEIQNEIYKTLAKNALLQLYNDKNH-----HFDELKQE 1189
Qy 1202 VN---KILKDKDEFFKEKENENNFIITASIYSLAKEDFITLLGDFKALKNKKGGEDYI 1257
Db 1190 FNIKNEEIPKTKAYDILRAK---NDLLVKKLYDLSDDFEYSWISTFKVL-NEKQSE--Y 1243
Qy 1258 SSLIK 1262
Db 1244 TTLK 1248
```

RESULT 8

```
OS0806 BORBU
ID OS0806 BORBU PRELIMINARY; PRT; 422 AA.
AC OS0806;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein BBK02.
GN OrderedLocusNames=BBK02;
```

```
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Plasmid lp36.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=96065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Caejens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Winn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi";
RL Nature 390:580-586 (1997).
DR EMBL; AE000788; AAC66148.1; -; Genomic_DNA.
DR PIR; D70251; D70251.
DR TIGR; BBK02; -.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 422 AA; 49766 MW; 2A87FB56008EFEAB CRC64;
```

Query Match 25.2%; Score 1648.5; DB 2; Length 422;

Best Local Similarity 75.7%; Pred. No. 5.4e-56;

Matches 315; Conservative 49; Mismatches 49; Indels 3; Gaps 2;

```
Qy 853 LKLIKEKGNLTLYLPSAIWNSSRIIRKHIFARYKLNIIYOFENKCRFKDVHSSPKFAI 912
Db 1 MKLVKPGGNLTLYLTPSSLSWSESGSKALRQHIFSNYKLNIIYOFENKQGFRAWTPGKFAI 60
Qy 913 FOLSNIEKSTSSPKAFWMIQSSDNILKEITRDLKSDKDAYKGIELNINQIKLSPIQES 972
Db 61 FOINSKEPTTKFRVKFVIQSSDNIMEKITSDLKAGENAYKGIELDIAQIKLSPIQES 120
Qy 973 IIEFKDNEEFTLINKMFSPKFSALGEGYIDFKKGLDPSIKNRKSLKECNKNLIFLYSGA 1032
Db 121 IIEFRDSEAEFTLVNKMFSRFDTLIQEYIDPREGL--NLTKYKALYKEYNEKFIPLYSGA 178
Qy 1033 NIHFNSRFPEDKDAKESKLLMIDKEDLEKVTLDNQYQTERVFYRAIASNTNERTMIS 1092
Db 179 NIHFNSRFPEDRAAKESKLLMIDKEDLEKVTLDNQYQAEVRYFYRAIASNTNERTMIS 238
Qy 1093 TILSPGNCVCNSIYINDEKTPISLYKKLFIISIFNSFVDFELLRFRVDSNVLKSCLYQCP 1152
Db 239 TILSPANCVCNSIYINTEEPISLYKKLFIISIFNSFVDFIIRFRVNSNVLKSCLYQCP 298
Qy 1153 MPQPEEKEILSNPLYLNLAKNTSLIAKNDPENPKYLLYLEYKFPDKEKNKILKDKED 1212
Db 299 MPQPEDEILNLSYLTAKNTSLIIVKNDPDNFKYLLYLEYFPFGKEEVDKMLNLDPKD 358
Qy 1213 EFPKEKENNNFIITASIYSLAKEDFITLLGDFKALKNKKGGEDYISSLIKGYDNYL 1268
Db 359 EFPKEKENNNFIIVASLYSLTKEDFVTLNDFVKCKN-KKGEDYISSLIKGYENYL 413
```

RESULT 9

```
OSXYI6 BORGA
ID OSXYI6 BORGA PRELIMINARY; PRT; 149 AA.
AC OSXYI6;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=EGP295;
OS Borrelia garinii PBi.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
```



```

Db 1120 GAVYPRRAAGDTFLOMLPDC-----PVAQIPGL--LGTLSNFAHDFAVRQKVG 1166
Qy 1142 NVLK-SCLYQCWPMPQPEEKEILSNPLYNLAKNTSLIAKNDPENFKYL----LYLEYFK 1196
Db 1167 THLYKYNVTRQPVLPF-----SAPTQRLAFITPRVLELTYYTA 1204
Qy 1197 PDKEKVNKILKDKEDBEFFKEKEN-----ENNFIASLYSLAKEDFIYLLGDFKALKNNK 1251
Db 1205 HDLAGFARDLGYDGPFPVWMDRFRWLRAELDALYFHLYGIPRDVDYIMETFPFIVRRKD 1264
Qy 1252 KGEDYISSLIKGYDNYLLNNKI 1273
Db 1265 EAQ-----YGYRTKNAI 1277

RESULT 11
Q7NGZ7 GLOVI
ID Q7NGZ7 GLOVI PRELIMINARY; PRT; 1442 AA.
AC Q7NGZ7;
DT 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE G12740 protein.
GN OrderedLocusNames=g12740;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Gloeobacteria; Gloeobacteriales; Gloeobacter.
OX NCBI_TaxID=33072;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PCC 7421;
RX MEDLINE=2977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimo S.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids."
RL DNA Res. 10:137-145(2003).
DR EMBL; BA000045; BAC90681.1; -; Genomic DNA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0008170; P:N-methyltransferase activity; IEA.
DR GO; GO:0008306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1442 AA; 163040 MW; C9F6G08F166E85C1 CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 3.4e-15;
Matches 332; Conservative 245; Mismatches 562; Indels 324; Gaps 69;
Qy 19 QLSKDFIKKENINK-----LKDFILIKNKLFSIDNTEANIES-LLKYIFEELNY 69
Db 24 ELEKLNREKLKSKATGTGAQALREFWDGYYRRKRLRELVGSKAIRNQVIDPILLPLRGY 83
Qy 70 S-VEQKAGOI-EGVESVDIILLFENDKDKASFNNKLEAKKNPEPIEDILLIAEYKR 127
Db 84 DRLEAAEPVQTRGLEAGHLAI-----SADGSKLRIWTD-----LDIDLDPAPKR 131
Qy 128 -PTFGDAKDKESDOLYRLNQYKHGYLSLNGKWRLY-----DKSKVLYGEKRY 180
Db 132 GAAYRF---SHURIAQ-----RVLLACSERVGLITNGVELRLLSDPARPDSVIIP----- 180
Qy 181 IBFNFXKIBEKEYEQEFVFLFYILIRKERYLKTSNVIEBKEQIAKEKEIIQKTLKEI 240
Db 181 IDANKRSREVPDSLR-----LLALASPAGKAVPGLVDKARLQQAQVTKDLRQAREA 235
Qy 241 LYERPDSDSVFKIANKIYDKFELSGKEITQHILASILEESIIFILRFFIAYIEDNIP 300
Db 236 V-----ELFQVRLDHPANRE-KLSEFADRARLAKELWREGLIITTYRLFILKJESSDDP 289

```

```

Qy 301 KKILQ--ENKLYRSSIS-----FRY--FFYDENTKK-KLEYKIIITITNLLDKGSDAIKF 350
Db 290 ARSFSFASTSLWRNSFSPSVALARYVMALDEGAETGQLLERGLRSLFRMFAEGLQCTEL 349
Qy 351 PV--FNGSLFSGEDKVKYLNNEGLLSIS--TEEILVKMLFEEKNIKDEKFKYKSRDPK 406
Db 350 NIRPLGSLFGAHTPILSD---LANGERAVAYLLDRLLTLTKGTVARERHYGPLDVE 406
Qy 407 SFGELYETLLBYDLRIADTTVHRI-----IEDG-VYLIR-----TEELENK-- 447
Db 407 DLGRVYEALLEPGIADPEMCLRRQKLEVVVPIDQGEKYRVTAAVAKNDSEDEAEEDA 466
Qy 448 -----KVNKIATYLGNIYLTSLDRKSKGAYYTPDLDITFPMWISSIEE 492
Db 467 EESQFEETSGRGKTKVEWIEAIAPGRPYLRV-GLGRKASGSYYTPHSFVFLVQETLGP 525
Qy 493 QLKTKSP-----LDIKIIDNSCGSHELISCLDYLTEKVMVYELDKFED----- 535
Db 526 QVARSPOSQPKPLEILKLVCDPAMSGHFLVACRFLGKL-YEARCLDELAMAAEK 584
Qy 536 -VKKELDEEYRVIIIESEY-----DVQDSISK-----EVLV 566
Db 585 RAESAEDQREAALEGAKLQORVADLPDPNDELLKYLPSRAPGLEIGLSQKAEAMC 644
Qy 567 KRMLLKXCIYQVDINPISVEITMLSLMINTFIFGTPLSFIHHIKAGNALLG-YTKD--- 622
Db 645 RRLIAVHCLYGVQDNPLAVELAKLSMTIESHAEGPLTFLDRLVLGDSLAGPFLKDLR 704
Qy 623 -----EPFDIVKKKFSFSLFKKRIKEIITILEDIYQKIKGINDTTKEDIKSK--- 672
Db 705 YPGSQKQMDLISLGLKEGFT---KALSNALRHVHDLSESV---GATLSEIAKQAAKT 757
Qy 673 KIKYEY-----ESKDIDNLRIFSLIKLYLSLSPDKSLME---FSD 711
Db 758 RLDKALEPFMIVAAAGGMLSGPECDSDAVARLAHAVASTSALPAD--INKEERLIRM 815
Qy 712 IASVISLIENLTGNKTSDEKIEKIRKLSYKFFHYGIE--PP-----DIQEGFDIVI 765
Db 816 IAKGMAQENL---TTIEVFQOKHIEKPLALPFDLTFVEVFPPHNSCNDYRNGFDVLL 871
Qy 766 GNPWEKTKPNETEFFSKHINRYKLGIKKEQNIKQETLSKDNH----- 809
Db 872 GNPPWSKSNL-----ELPEY--MGAVDPFRL--EVTSSDERIDLESYKSSANWAW 919
Qy 810 -PLSIEYNEEKNSIITANNIYKDFPKCPTSGDGNLFRYFVTFNLKLIKEGNLYLVP 868
Db 920 LLASLEEDBMCKIISLNPDAQSDLVGFGS-GHADIFASFDVDRFLSLRVNGLFAMVLP 978
Qy 869 AIWNSSSRILRKHIFARYKLYNYIQPENKRFKDVHSSPKP--AIFOLSNIKESTSPK 926
Db 979 SLHVSQSLQRLRRKLLAKAEILSVFSPENKYSLEFHKSWKFTPVIIIRNSEPKEK-KRFP 1037
Qy 927 AKFMQSSDNLKKEITRDLKSDKDAYKGIELNINQIKLSPIOESIIEFKDNEFTLIN 986
Db 1038 AQFYLHDSWLFSA-----NKDPYYLYDPNIEVMD-----SESLI-----FQBYTSVS 1082
Qy 987 KMFKFSALGEGYIDFKKGLDP---SIKN-----RKSLLKEC-----NNKNLIFY 1029
Db 1083 DMKCAETIFGNS-LSWKYSENTYWSIRRELNATDNRWIEYSKPCGRTWSKDEALIH 1141
Qy 1030 SGANTHQFNRFEDKAKESKLLWIDKEDLEKVLTKDNQYQTERVF-----YRA 1080
Db 1142 QPGTIHQYTD-----LWTGAKTL--TIPFNLYDRPVLELSRYFAAYRM 1185
Qy 1081 IASNTNERTMLST-LSPGNCYCVNSIYIN--DEKTPISLYKKLFIISFNSFVDFLRR 1137
Db 1186 SARATDERTSITILTPTGT--ATNSLPTEGCPQKRPNRL--SLGAVAVCNNSFVDFWYLR 1242
Qy 1138 FVDSNVLAKSLCYQCPMPQPEEKEILSNPLYNLAKNTSLIAKNDP--ENPKYLLYLEVF 1195
Db 1243 RVGSKTISKFI-----MDNTPISDLASQN---NLCVHSSLRVLSNHPGVTLWHEQVCDEW 1295

```


RA	Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,	
RA	Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,	
RA	Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,	
RA	Utterbach T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,	
RA	Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,	
RA	Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;	
RT	"Complete genome sequence of the methanogenic archaeon, Methanococcus	
RT	jannaschii";	
RT	Science 273:1058-1073 (1996).	
RL	-1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-	
RL	adenosyl-L-homocysteine + DNA 6-methylaminopurine.	
CC	-1- SIMILARITY: Belongs to the N4/N6-methyltransferase family.	
CC	-----	
CC	This Swiss-Prot entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use as long as its content is in no way modified and this statement is not	
CC	removed.	
CC	-----	
CC	EMBL; L77119; AAC37060.1; -; Genomic_DNA.	
DR	PIR; B64516; B64516.	
DR	TIGR; MJEC502; -	
DR	InterPro; IPR011639; Eco57I.	
DR	InterPro; IPR002296; N12N6_mtfase.	
DR	InterPro; IPR002052; N6_Mtase.	
DR	Pfam; PF07669; Eco57I; 1.	
DR	PRINTS; PR00507; N12N6MTFRASE.	
DR	PROSITE; PS00092; N6_MTASE; 1.	
KW	Complete proteome; Hypothetical protein; Methyltransferase; Plasmid;	
KW	Transferase.	
KW	SEQUENCE 1181 AA; 138439 MW; 9F961D8F6C6A4BDD CRC64;	
QY	Query Match	
QY	Best Local Similarity 8.1%; Score 526.5; DB 1; Length 1181;	
QY	Matches 322; Conservative 212; Mismatches 399; Indels 509; Gaps 76;	
QY	-----	
QY	25 IKKENINKLKDFFILKKNKLFSDDD-NSTEANIESLLK--YIFEBLNSVVEOOKAGQI- 79	
Db	15 IPKEVYNKKWDFIKLVGNILSIKQIPELEGKFKELLKKGRYNFKSDDF-----GGQLP 68	
QY	80 -----EGVESRVDIILLFENDKDKASFNKKL-KEAKKNNPEIP-----IEDILLIAEV 125	
Db	69 EPFTROKVIETPILEFLGYE-----FTSEISKKSPLGDRKIPDYRSVFNKEILTEAE- 120	
QY	126 KRPTFS-FPAKO-----KLKESEDQLYRVLNOYKQHYGLISNGKWRLYDKSKVLYGEKRY 180	
Db	121 --PLGSDLNKKDSGIIHQVKE-----WLIITSYQVDTGIATNGLEWLLVHDDTIKEIRTL 173	
QY	181 IEFNFKKIEE-----KEEYKQEW---FVLFYTLIRKERYLKTSNVIEKEQIA-KEK 230	
Db	174 KELNLSIPEYVLNKKDKDLENELQVSEFYFCFSKE-YIE-EVIEVATKNIKHKKE 230	
QY	231 EIIQTKLKEILYRPPDSIV--FKIAKNIYDKFELSGKEI-TOHILASILEE----- 280	
Db	231 EITNEFYKSFV-----KLVGFGEDEVKQVKDKSSKDKGFKCKLYNCIEAPPNTSEL 284	
QY	281 -----SIITFLRIFFIAYIED-----NDIFPKILQENKLVRSISIRPYF-----FYD- 322	
Db	285 DKKKFVALLMNRILFIKFLDEKGIQVPRDLRRYYEDYK--KSNVLINYYDAYLKPFLPYEV 342	
QY	323 -----ENTKKKLEYKKIITIFNLLDKGSDAIKFPVNGGFLGEDSKVKYLLNKGLLS 373	
Db	343 LNTPEDERKENIRTPYKDI-----PYLNGGLFRSNVNP---NELSFT 383	
QY	374 ISIEEILVQMLPFESK---NIKDEKPKVYKRLDPKPSFGELYETLLEYLDIRIADTTVHRI 430	
Db	384 IKD-NEIIGEVINFLERYKFTLTSTSGSEVEVLNPDILGYVE----- 425	
QY	431 IEDGVVLIETEELENKYNKIATYKGNILYTSRSLDRKSKGAYYTPDDLTDFFWISSI 490	
Db	426 -----KLINILAE--KG-----QKGLGAYYTPDEITSYIAKNTI 457	
QY	491 E----EQLK-----TKSPLD-----IKIIDNSCG 510	

Db 458 BPVVVERFKBIKNWKINDINFSTDEILNEDSKIAENKHILRAFLDELDIRILDPVAVG 517
QY 511 SGHFLISCLDYLTKEVWVYELDKFEDVKKELDEYRVIIIEESEEYVDQDSISKELVLKML 570
Db 518 SGHFLISALKELLQ-----IKKRI---YLLREMDIY-----KE---KLGI 553
QY 571 LKXCIYGVNDINPISVEITWLSWI-----NTFIFGTP-----LSFIEHHIKAGNALLGYTK 621
Db 554 ILNLYGVVDIDDIJAEIAKRLALALLENLOVEALKRGEVLLPNIEYVNRVCGNSLAVGWD 613
QY 622 DEFFDVKKKPESGSLFKRIKEIITILEDIYQIKGINDTTKEDIKSKKIYKEYEES 681
Db 614 ENL-----KQISISLCONVR---IMCVLEGLI-----INAHNSEERKKLKA-KELLEK 659
QY 682 KD---IDNLRIFSLI-KLYSLSFDKSLNM-----EFSDIASVISLJENIL----- 723
Db 660 RDGYVLDNVEAYHLLYEVYRSHGLKANLLKELDEIRD-----SIYESVTPPAYPAIY 714
QY 724 ---GN-----KTSSEKKEKIRKLSYYKFFHYGIBFPDI--OEGFDIVIGNPPWEKT 773
Db 715 QNGNKKNGKKSKKRNPRVEFEKL---KPFHWKIDFGWIIKEEGFDVIIIGNPPYGNL 770
QY 774 KENETEFPSKHIPNRYKLGIKEQNIQKQILSKONHPLSIEYNEEKNSIIAINNIYKEDF 833
Db 771 -LSPT-----KEIMKRRDTP-----EFD- 788
QY 834 KCFTSGGDPNLPFYVTFNMLKLIKEGNLTLYLPSAIWNNESSRILRKHFARY---KLN 890
Db 789 -----IPVTFIVHSSKLLKNEGVLGFIIPSSFGTGVYRSLRKLFTYKMLCKLI 838
QY 891 Y-----IYQFNKKRPFKDVHSSPKFAIFQLSNIKE 920
Db 839 YLPFDVFGAYVDCNIIILHKPKPKSEDLVLIYAPKTK----- 878
QY 921 STSSPKAFEMIQSSDNILKEITRDLKDSKDAY-KGIELNINOIKKLSPIQESIIEPKDN 979
Db 879 -----KISPEFKNDLFIEYSKILNDPKRIFPKSPIIYIILDKIKQNCRESLYLEDL 931
QY 980 BEFTLINKWFSKF---SALGEGYIDFKG-----LPSIKNRKSLKCECNKNLIPLY 1029
Db 932 TEST-IGILASKYKESDKKENEYIPLYEGNVRYETKLNKYVDVFSKHKNKELINLF 990
QY 1030 SGANIHQFNRPFDKAKESKLLWIDKEDLEKVLTKDNQYQTERVFVRAIASNTNERT 1089
Db 991 -----MSPEKIFIRIV-NRQDRI 1008
QY 1090 MISTLSPGNCYCVNS-----IYNDEKTPISLYKKLFIISFNSFVDFDLRRFVDSN--V 1143
Db 1009 MASY---GNIEGWKKDLVVFVKPDTPINYF---YLLGIILNSELISYI---YIGKSAIA 1059
QY 1144 LKSCLYQCPMPQEPKEILSNLYLNAKNTSLLAKNDPENFKYLLVLEYKFDKKN 1203
Db 1060 LKDDFRQTTLELRPLIV-----IPKNKIINA-----LTQLSKURFELND-- 1101
QY 1204 KILKLDKEDEFFKEKENENNFIASLYSLAKEDFITLLGDFKALKNKKGGEDIYISSLING 1263
Db 1102 ---KLNENDRIFLE-----NIIDSLVYGIYFQDLIP-----KEELNEICNE--INGIICK 1146
QY 1264 YD 1265
Db 1147 YD 1148

Search completed: January 24, 2006, 19:56:56
Job time : 280.4 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 78.7244 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-16

Perfect score: 6538

Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIKYDNYLLNNKIFYHK 1277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5_COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6_COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H_COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE_COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	394	6.0	1218	US-09-605-703B-600	Sequence 600, App
2	382	5.8	1009	US-09-605-703B-146-4	Sequence 4, Appli
3	316	4.8	1602	US-09-269-874A-7	Sequence 7, Appli
4	316	4.8	1621	US-09-269-874A-5	Sequence 5, Appli
5	316	4.8	1639	US-09-269-874A-3	Sequence 3, Appli
6	315	4.8	2184	US-09-417-485B-6	Sequence 6, Appli
7	309	4.7	622	US-09-605-703B-604	Sequence 604, App
8	285.5	4.4	2662	US-09-595-684B-31	Sequence 31, Appli
9	285.5	4.4	2663	US-09-538-092-1252	Sequence 1252, Ap
10	280	4.3	990	US-08-392-625-20	Sequence 20, Appl
11	280	4.3	990	US-08-466-961A-20	Sequence 20, Appl
12	278	4.3	990	US-08-645-193B-15	Sequence 15, Appl
13	276.5	4.2	1007	US-09-957-005-9	Sequence 9, Appli
14	269	4.1	2710	US-08-480-604A-6	Sequence 6, Appli
15	269	4.1	2710	US-08-405-496A-6	Sequence 6, Appli
16	269	4.1	2710	US-08-915-136-6	Sequence 6, Appli
17	269	4.1	2710	US-08-957-310-6	Sequence 6, Appli
18	269	4.1	2710	US-10-011-366-6	Sequence 6, Appli
19	269	4.1	2710	US-09-084-517-6	Sequence 6, Appli
20	265.5	4.1	1467	US-09-134-000C-6740	Sequence 6740, Ap
21	262.5	4.0	2777	US-09-543-681A-6124	Sequence 6124, Ap
22	261.5	4.0	1786	US-08-973-462-8	Sequence 8, Appli
23	259	4.0	1068	US-09-966-997-11	Sequence 11, Appl
24	255	3.9	1477	US-09-830-230A-414	Sequence 414, App
25	255	3.9	1494	US-09-830-230A-413	Sequence 413, App
26	252	3.9	3169	US-09-453-702B-257	Sequence 257, App
27	252	3.9	3169	US-10-114-170-257	Sequence 257, App

28 250 3.8 1119 2 US-09-830-230A-321 Sequence 321, App
29 247 3.8 1381 2 US-09-662-254B-20 Sequence 20, Appl
30 246 3.8 1087 2 US-09-830-230A-322 Sequence 322, App
31 245 3.7 3878 2 US-09-914-259-11 Sequence 11, Appli
32 239 3.7 648 2 US-10-338-898-2 Sequence 2, Appli
33 238 3.6 905 2 US-09-248-796A-16333 Sequence 16333, A
34 237 3.6 1173 2 US-09-248-796A-19313 Sequence 19313, A
35 236.5 3.6 1290 2 US-10-360-101-220 Sequence 220, App
36 236.5 3.6 2733 2 US-09-949-016-11433 Sequence 11433, A
37 236.5 3.6 3259 2 US-09-949-016-6507 Sequence 6507, Ap
38 235.5 3.6 3135 1 US-08-323-170B-2 Sequence 2, Appli
39 235.5 3.6 3135 2 US-08-954-441-2 Sequence 2, Appli
40 233.5 3.6 618 2 US-09-299-378-4 Sequence 4, Appli
41 233.5 3.6 1010 2 US-09-134-001C-5178 Sequence 5178, Ap
42 233 3.6 1257 2 US-09-245-928A-18 Sequence 18, Appli
43 232 3.5 1257 2 US-08-947-823-5 Sequence 5, Appli
44 229.5 3.5 10182 2 US-09-134-001C-3159 Sequence 3159, Ap
45 227 3.5 1078 2 US-09-248-796A-20284 Sequence 20284, A

ALIGNMENTS

RESULT 1

US-09-605-703B-600
; Sequence 600, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 600
; LENGTH: 1218
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-600

Query Match 6.0%; Score 394; DB 2; Length 1218;
Best Local Similarity 19.0%; Pred. No. 6e-17;
Matches 211; Conservative 189; Mismatches 379; Indels 332; Gaps 42;
QY 331 YKTIITFNLLDK-----GSDAIKFPVNGGLFSDKVKYLNNEGLLSI 374
DB 27 YDSLQLLFTLVDRKHHTDPAKAVDAEAGEEGLHFNLSADLFLPAATELIDRVG-LSN 85
QY 375 SRIEELVKKMLFEEKNIKDFVKYSLDDPKSFGELVETLLEYDLRIA-----DTTVHR 429
DB 86 EALNKVLENLLLSRVQSGKDRGFISYATLGVTELGOVLEGLMSYTGFIQAEDLFEVAPHG 145
QY 430 IIEDGVYL-----IRTEEELENKKVANKI-ATYLGKNIYLTSSRLDRKSGA 474
DB 146 KADKGSWMLPVSKADEVPADSFIEVDQAPGGGVKVRKRPGRGSVFRRQSRDRERSAS 205
QY 475 YTPDDLTDFWISSIEEQLTKY-----SPDKIKIINDNSCGSGHFLISCLDYLTKEVWY 528
DB 206 FYTPQVLTSTFTQATIEELQASKRITTANDVLSLTICEPAMGSGAPAVEAVRQLAE-LYL 264
QY 529 ELDKEDVKKELDEEYRVIIIESEEDVDQDSISKEL-VLKRMLLKCCIYGVNDINPISVEI 587
DB 265 EL-ROEELEQQIPAEOR-----AKELQKVKAHIALHQVYGVVDLNSTAVEL 308


```
QY 947 DSKDD--AYKGIENINQIKLSPIQESIIIEPKDNBEFTLINOMFSKFSALGEGYIDPKK 1004
Db 670 DSBENTCVSSBKYNSDPWIFLSPETEAV-----FTKFTB--AQFEKLG-ITDISV 718
QY 1005 GL-----DPSIKNRKSLKE-----C-----NNKGLI 1026
Db 719 GLQTSADKIYIFIPENETSITYIFNYKGRYEIEKSIKCCPAIYDLSFGSPESIQNAQMI 778
QY 1027 FLYSGANIHFOPGRFFEDKAKESKLLWIDKEDLEKVLTKONQOTERVFYRAIASNTN 1086
Db 779 FPY---EIRDEEAYLLEETELENDDYPLAWNYLNEFEKLEKRSIQGRNPKWYQGRSQSL 835
QY 1087 ER-----TWISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISIFNSVFEDFELRR 1137
Db 836 SKFHDKEKLIWTVLATKPPVVLDRNLLFTGGNGP---YGLINOSIYSLHYF----- 886
QY 1138 FVDSNVLSKLYQCPMPQPEBEILSNPLYNLAKNTSLIIAKNDPENFKYLLYLEYKFK 1197
Db 887 -----LGLSHPVIESWVKARA-----SEFRG---SYSH 913
QY 1198 DKEKVNI-----LKLKDEBFFK-----EK-----ENENN----- 1223
Db 914 GKQFIEKIPIRKIDFDDQDEVDKYNTVTTVEKLIITTDRIKESNGPRRRMLRRILDAL 973
QY 1224 -----FIASLYSLAKEDFITILGD 1243
Db 974 SNQLIQVINELYNDEEYTTVLND 998

RESULT 3
US-09-269-874A-7
; Sequence 7, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSPI
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269, 874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1602
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-7

Query Match 4.8%; Score 316; DB 2; Length 1602;
Best Local Similarity 19.7%; Pred. No. 9e-12;
Matches 295; Conservative 229; Mismatches 481; Indels 490; Gaps 73;

QY 2 KTNDI-----VKTNNPNISLYKQLSKDFIK-----KENINKLDPFILLIKNKL 45
Db 172 KLVDCANDYCOIPFNKLIRANELDLVLLKLVFGYKPLDNKONVGKMDY--IKNK-- 227
QY 46 SIDDNSTEANIESLLKYIFELNYSVEQQKAGQIGEVESRVDILLFENDKKKASFNKKL 105
Db 228 -----KTNIENINELI-----EESKTDKNKNATKEBEKK-----LYQAQYDLSYNNKQLE 274
QY 106 EA-----KKNNEPIEDIL-IIAEVR-----PTSFDAKDKLES 141
Db 275 EAHNLISVLEKRIIDLTKNEN-----IKELDKINEIKNPPANGSNTPTNLLDRNKKIEEH 331
QY 142 E-----DOLYR-----YLNQYOKHYGILS-----NGKWR 166
Db 332 EKSIKEIAKTIKNIDSLFTDPLEYLYLREKKNIDISAKVTKESTEPNEVPNGVTYP 391
QY 167 L-YDKSKVLYGKRYI-----EFNFXKIEEYKEQWFLVFIYLIRKERYLKTNSVIE 220
```

```
Db 392 LSYNDINNALNELNSFGDLINPPDYTKBPSKNIYTDNE-----RKFFINEIKEKIK 442
QY 221 VKEQIAKEKEIIQ---KTLKEILYERPDSDIVFKIAKNIYDKEF-----KLSGK 267
Db 443 IEKKKIESDKSYEDRSKSLNDITKEYE-----KLLNEIYDSKFNNNDILTTFKMMGK 496
QY 268 -----EITOHILASILEESIIFILRIP-FTAIYED-----NDIFPKILOENKLYRSSIS 315
Db 497 RYSYKVEKLTHTNTFASYENSKHNLEKLTAKALYMEDYSLRNTIVVEKELKYKLNLSKIE 556
QY 316 FRYFFVDENTKK--KLEYKKLIITIFNLDDG---SDALKPPVFNGLFSEDKVKYLYNN 368
Db 557 NEIETIVENIKDDEBQLEFKKTKDENKPEDEKILEVSDIVKVOV----- 600
QY 369 EGLLSISEIEILVLMKLFEEKNIDEKFKYSRLDPKSGFELYETLEIDY---LRIADT 425
Db 601 QKVLLANKIDELKKTQLILKNVELKHNHVPNSYKQENQEPYLYLVKKEIDKLVFMP 660
QY 426 TVHRIITDGVYLRTEEBELENKKNKIATYLVKNIYLTSLRSLDRKSGAGYTTDPDDLTFDM 485
Db 661 KVESLINEBKNIKTGQSDNSEPS-----TEGEITQATTKPGQOAGSALEGDS----- 710
QY 486 VISSIEQLKTKSPLDIKI-----IDNSCGSGHPLISCLDYLTEKVM----- 527
Db 711 VQAQEQKQAQPPVPVPPEAKAQVTPPAPVNNKTEN---VSKLDYL-EKLYBFLNT 765
QY 528 -----YELDKPEDVKKELDEYRVIIEESEYDVQDSISKELVLKRLMLKXCIYGVGIN 581
Db 766 SYICHKYILVSHSTWNEKILQYKITKEESKSSCDPLD-----LLFNQI-N 812
QY 582 PISVEITMLSLMINTPIGTPLS--FIEHHIKAGNALLGYTKDEFFDIVKKKESGESL 639
Db 813 NIPVMYSMPDSLANS-----LSQLFMEIYKEMVNCVLYKLDN--DKIKNLEEA----- 860
QY 640 KKEIKIITILEDIYQIK-----GINDTTKE-----DIEKSKKIY 675
Db 861 KKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLSNLSKLFENILSLGKNKIY 920
QY 676 KEYEESKIDNL--RIIFSLIKLYSLSPKSLNMFSDIASVISLIENTILGNKTSSEDK- 732
Db 921 QELIGOKSSENFYEKILKDSDFYNESFTNFVKSADDINSL-----NDESKKKL 971
QY 733 -EKIEKIRKLSYYKFFHYGIBFPDIQEGFDIVIGNPPWEKTKFNTEFPFSKH--IPNYR 789
Db 972 EEDINKLKK-----TLQLSFDL-----YNYKYLKLERLFDKKTGVGY- 1009
QY 790 KLGIKEQNIQKQELSKDNHPLSIEYNEEKNISIIANNIYKPFKCFCTSGDNLPRYFV 849
Db 1010 KMQIKKLTLLKEQLESKLN-----SLNN----- 1032
QY 850 TPNLKLKEKGNLTLYLPSAIWNNESSRILRKHIFARYKLYNIYQFENKKRFDKDVHSSPK 909
Db 1033 -----PKVLQNFV-----FNNKKKEAIEAET-- 1055
QY 910 FAIFQLSNIKESTSPFAKFMIQSSDNILKEITRDLKSDKDDAYKGIELNIN-QIKKLS 968
Db 1056 -----ENT-----LENTKILKCH-----YKGLVKYNGESSPLKT 1085
QY 969 IQESIIEFKDN-----BEFTLINMFKSFALGEGYIDFKK-----GLDPSIKRKS 1016
Db 1086 LSEESIQTEDNVASLENFVLSKLEGL-----KDNLNLEKKKLSYLSGSLHLHLIABLKEV 1141
QY 1017 LKECNKNKILFLYSGANIHFQNSRPFEDKAKESKLLWIDKEDLEKVLTKD-----NOY 1071
Db 1142 IK---NKN-----YTGNSPSENNT---DVNNALSYKFFLPEGTDVATVVSSESDTLEQS 1191
QY 1072 QTERVFYRAIASNTNERTMISTLSPGNCYCVNSIYINDEKTPISLYKKLFIISIFNSVF 1131
Db 1192 QPKKPASTHVGAESNTIT-----TSQNVDDDEVDV-----IIVIFGSESE 1232
QY 1132 DF-----LLRRFVDSNVLK-----SCLYQCPMP---QPEKEILSNPLYNLAK 1172
```

Db 1233 DYDDLQGVTVGEAVTPSVIDNILSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVV 1292
Qy 1173 NTSLLAKNDPENFKVLLYLEYKFP--DKEKVNKILKLDKEDEF--FKEKENENNFIILASLY 1230
Db 1293 KDILNRFNKNFKNVLESDDLIPYKDLTSSNVVVK---DPYKFLNKEKRDKFL--SSY 1346
Qy 1231 SLAKEDFIT-----LLGDFKALKNNKKE-----DYISSLIKGYDNYL--LNN 1271
Db 1347 NYIKSDITDINFANDVLGYKILSEKYSKDLSDLSIKKYINDKQENEKYLPFLNN 1401

RESULT 4
US-09-269-874A-5
; Sequence 5, Application US/09269874A
; Patent No. 693130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSPI
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1621
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-5

Query Match 4.8%; Score 316; DB 2; Length 1621;
Best Local Similarity 19.7%; Pred. No. 9.1e-12;
Matches 295; Conservative 229; Mismatches 481; Indels 490; Gaps 73;

Qy 2 KTNDI-----VKTNNPNIISLYKQISKPIK-----KENINKLDPFILIKNKLP 45
Db 191 KLANDVCANDYQIPFNLKIRANELDVLKLVFGYRKPLDNIKNVGMEDY--IKNK-- 246
Qy 46 SIDDNSTEANTIESLKKYIFELNYSVQKAGQLEGVESVDILLFENDKDKASFNNK 105
Db 247 -----KTNIENELI-----BESKTDIKNKATKEBEKK-----LYOQYDLSLYNKQLE 293
Qy 106 EA-----KKNNEPIPIEDIL--IIAEVVR-----PTFSFDAKDKLES 141
Db 294 EAHNLISVLEKRIIDLTKNEN---IKELLDKINEIKNPPANSNGTPTNLLDKNKKIEEH 350
Qy 142 E-----DQLYR-----YLNQYQKHGYLS-----NGKWR 166
Db 351 EKEIKEIAKTIFKNIDSFTDPLBLEYLRKKNKIDISAKVETKESTEPNEYPNGVTP 410
Qy 167 L-YDKSKVLYGEKYI-----EFNFKIEEKEEYKQEWFLFYILIRKERYLAKTSNVIE 220
Db 411 LSYNDIINNALNELNSFGDLINPFDYTKEPSKNYITDNE-----RKKFINEIKKIK 461
Qy 221 VEKEQIAKEKEIIQ---KTLKEILYLRPDDSVIFKIAKNIYDKEF-----KL-SGK 267
Db 462 IEKKIESDKSYEDRSKSLNDITKEYE-----KLNLEIYDSKFNNDILTNPEKMWGK 515
Qy 268 -----EITOHILASILESIIIFILIP-FIAYED-----NDIFPKILQENKULYRSIS 315
Db 516 RYSYKVEKLTHHTFASVENSKHNLKLTALKATKMTMEDYSLRNIIVVEKELTYKQLISKIE 575
Qy 316 FRYFFYDENTKK--KLEYKIIITFNLLDKG---SDAIKFPVFNGLFSDEKVKVLYNN 368
Db 576 NEIETLVENIKDEQLFEKKTIDENKPDKEILEVSDIVKQV----- 619
Qy 369 EGLLSISIEBILVKMLPFEEKNINDEKFKVYSLDPKPSFGELYETLLEYD---LRIADT 425
Db 620 QKVLLMKNIDELKKTQLTLKVNKLKHNHVNYSYKQENKQBPYLLIVLKBIDKLVFMP 679

Qy 426 TVHRIIEGCVYLIRTEEBLENKVNKIATYTLKGNITVLTSRSLDRKKSGAYTTPDDLDFM 485
Db 680 KYESLINEKNIKTEGQSDNSEPS-----TEGEITGQATTKPGQAGSALSGDS----- 729
Qy 486 VISSIEBQKTSKPSIDIKI-----IDNSCGSGHFLISCLDYLTBKVM----- 527
Db 730 VQAQAEQKQKQAPPPVPEAKAQVTPPAPVNNKTEN---VSKLDYL-EKLYEFLNT 784
Qy 528 -----YELDKPEDVKKDELDEBYRVIIESEBYDVQDSISKELVLKMLLKXCIYGVGIN 581
Db 785 SYICHKYILVSHSTWNEKILKQYKITKEESKLSGCDPLD-----LLFNIQ-N 831
Qy 582 PISVEITMLSLMINTFIFGTPLS--FIEHHIKAGNALLGYTKDEFFDIIVKKKPFESGSLF 639
Db 832 NIPVMYSWFDLSNNS-----LSQLFMEIYEKEMVCNLYKLKON--DKIKNLLSEA--- 879
Qy 640 KKRIRKIIITLBDIYQIK-----GINDTTKE-----DIEKSKKIY 675
Db 880 KKVSTSVKTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFPENILSLGKNKIY 939
Qy 676 KEYEESKIDNL--RIIFSLIKLYSLSPDKSLNMEFSDIASVISLIENILGNKTSSEDK- 732
Db 940 QELIGQKSENFEYKILKDSDTFFYNESTFNFKSKADDINSL-----NDESRRKKL 990
Qy 733 -EKIEKIRKLSYVYKFFHYGIEFPDIOEGFDIVIGNPPWEKTKFNETEFFSKH--IPNYR 789
Db 991 EEDINKLKK-----TLQLSFDL-----YNYKYLKLERLFDKKTGVKY- 1028
Qy 790 KLGIKEQNIIOKILSKONHPLSIEYNEEKNSIIAINNIYKDFKCFTSGGDNLFRYFV 849
Db 1029 KMQIKKLTLLKEQLESKLN-----SLNN----- 1051
Qy 850 TFWNLKILKEGNLTVLVPISAINWESSRILRKHIFARYKLNLYIYQENKPKKPDVHVSFX 909
Db 1052 -----PKHVLQNFVS-----FPNKKEABIAET-- 1074
Qy 910 FAIFQLSNIKESTSSPKAFMIQSSDNILKEITRDLKDSKDADYKAGIELNIN--QIKKLS 968
Db 1075 -----ENT-----LENTKILLKH-----YKGLVKYNGESSPLKT 1104
Qy 969 IOESIIEPKDN-----BEFTLINRKFSGALGEGYIDFKK-----GLDPSIKNRKSL 1016
Db 1105 LSEESIOTEDNYASLENFKVLSKLEGL---KONLNLEKKKLSYLSLSSGLHLIAELKEV 1160
Qy 1017 LKECNKNLIFYSGANIHOFSRPFEDKAKESSKLLWIDKEDLEKVLTKD-----NOY 1071
Db 1161 IK---NKN---YTGNSPSENNT---DVNNALSESYYKKFLPEGTDVATVSVESGSDTLEQS 1210
Qy 1072 QTERVFPYRAIASNTNERTMISTLSPGNCYCVNSIYINDEKTPISLYKCLPIIIFNSFVF 1131
Db 1211 QPKKPASTHVGAESNTIT-----TSQNVDDDEVVV-----IIVIFGSEEE 1251
Qy 1132 DF-----LLRRFVDSNVLK-----SCLYQCPMP---QPEKEILSNFLYLNLA 1172
Db 1252 DYDDLQGVTVGEAVTPSVIDNILSKIENEYEVLYLPLAGVYRSLLKQLENNVMTFNVV 1311
Qy 1173 NTSLLAKNDPENFKVLLYLEYKFP--DKEKVNKILKLDKEDEF--FKEKENENNFIILASLY 1230
Db 1312 KDILNRFNKNFKNVLESDDLIPYKDLTSSNVVVK---DPYKFLNKEKRDKFL--SSY 1365
Qy 1231 SLAKEDFIT-----LLGDFKALKNNKKE-----DYISSLIKGYDNYL--LNN 1271
Db 1366 NYIKSDITDINFANDVLGYKILSEKYSKDLSDLSIKKYINDKQENEKYLPFLNN 1420

RESULT 5
US-09-269-874A-3
; Sequence 3, Application US/09269874A
; Patent No. 693130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a

; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSP1
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-3

Query Match 4.8%; Score 316; DB 2; Length 1639;
Best Local Similarity 19.7%; Pred. No. 9,2e-12;
Matches 295; Conservative 229; Mismatches 481; Indels 490; Gaps 73;
QY 2 KTDNI-----VKTNNPISLYKQSKDFIK-----KENINKLQDFILIRNKLF 45
Db 191 KLDVDCANDYCOIPFNLKIRANELDLVKLVFGYRKPLDNKONVGKMDY--IKKTK-- 246
QY 46 SIDDNSTANIESLLKYIFELNYSVEQKAGQIEGVESRDVILLFENDKXKASFNKXK 105
Db 247 -----KTIENINELI-----EESKTDKNKNATKEBEKK-----LYQAQYDLSIYNKQLE 293
QY 106 EA-----KKNNEPIPIEDIL-IAEVRK-----PTFSFDAKDKLKE 141
Db 294 EAHNLISVLEKRIDTLKKNEN---IKELDKINEIKNPPNPANSNTPTLLDKNKIEEH 350
QY 142 E-----DQLYR-----YLNQYQKHGILS-----NGKWR 166
Db 351 EKEIKEIAKTIFKNIDSLFDPLELYLREKKNIDISAKVETKESTEPNEPVGVTY 410
QY 167 L-YDKSKVLYGKRYI-----ENFYKIEBESEYKQEWFLYILIRKERYLKTNSVIE 220
Db 411 LSYNDINNALNELNSFGDLNPFDTYKPSKNTYTONE-----RKKFINEIKEIK 461
QY 221 VEKEQIAKEBIIQ---KTLKEILYERPDSDIVFKIAKNYDKEF-----KLSGK 267
Db 462 IEKKIESDKSVEDRSKSLNDITKEYE-----KLLNEIYDSKFNNDILTWFERNMGK 515
QY 268 -----EITQHILASILEBSIIFILRIF-FIAYIED-----NDIFKKILOENKLYRSSIS 315
Db 516 RYSYKVEKLTHNTFASYSNKENLEKLTAKALYMEDYSLRNIVVEKELAYYKNLISKIE 575
QY 316 FRIFFYDENTK---KLEYKTIITIFNLLDKG-----SDAIKPPVFNGLFSEDKVKYVLNN 368
Db 576 NEIETLVNKKDBEQLEFKKTKDENKPKDEKILEVSDIVKQV----- 619
QY 369 EGLLSISEIEILVKMLFPEKNIKDKFVKYKSLDPKSGFELYETILEYD---LRIADT 425
Db 620 QKVLMMNKIDELKKTQILKNVLEKNHVPNSKQENKQEPYLLVLRKEIDKLKVFMP 679
QY 426 TVHRIEDGVYLIRTEBELENKVNKIATYLGKNIYLTSLDRKSKSGAYTTPDDLTDFM 485
Db 680 KVESLINEEKNKITEGQSDNSPES-----TEGEITQATTKGQQAQSALEGDS----- 729
QY 486 VISSIEQLTKPLDIFI-----IDNSCGSHFLISCLDYLTEKVM----- 527
Db 730 VQAQAQEQKQAPVPVPVPEAKAQAQVTPPAPVNNKTEN---VSKLDYL-EKLYEPLNT 784
QY 528 -----YELDKPFDVKELDEEVRVIERSEEDVDQDSISKELVLKMLLXCIYGVGIN 581
Db 785 SYICHKYLVSHTMNEKILQKVKITKEESKLSKSCDPLD-----LLFNQI-N 831
QY 582 PISVEITMLSLWNTFFIGTPLS--PIERHIKAGNALLGYTKDBFFDIVKVKPESGSLF 639
Db 832 NIPWYMSFSLNNS-----LSQLFWEIYEKEMVCNLYKLDKN--DKIKNLILEEA----- 879
QY 640 KGRKEIITILEDIYQKIK-----GINDTTKE-----DIEKSKKIY 675

Db 880 KKVSTSVKTLSSSSMQPLSLTTPQDKPEVSANDTSHSTNLSNLSKLPEMILSGKNKIY 939
QY 676 KEYBESKDNLN--RIIFSLIKLYSLSPDKSLNMFSDIASVISLIENILGNKTSBEDK- 732
Db 940 QELIGQKSSSENFVEKILKSDTFFYNESFTNFVKSKADDINSL-----NDESKRKKL 990
QY 733 -EKIEKIRKLSYVYKFFHYGIEFPDIOEGFDIVGNPPWEKTKFNETEPFSKH--IPNYR 789
Db 991 EEDINKLKK-----TLQLSFDL-----YNNKYKLKERLFDKKTGVGY- 1028
QY 790 KLGIEQNIIOEILSKDNHPLSIEYNEEKNSIIAINNIVKDFKCFKTSGGDNLFRYFV 849
Db 1029 KMQIKKLTLLKEQLESKLN-----SLNN----- 1051
QY 850 TPNLKLIEKGNLTLYLPSAINESSRILRKHIFARYKLNLYIYQENKGRFQDVHSSFK 909
Db 1052 -----PKHVLQNFVS-----PFNNKKEABIAET-- 1074
QY 910 FAIFQLSNIKESTSTSPKAFMIQSSDNIKEITRDLKDSKDDAYKGIELNIN-OIKKLSP 968
Db 1075 -----ENT-----LENTKILLKH-----YKGLVYKNGESSPLKT 1104
QY 969 IOESIIEPKDN---EEFTLINKMFSKFSALGEGYIDFKK-----GLDPSIKNKRKSL 1016
Db 1105 LSEESIQTEDNYSASLENFKVLSKLEGL-----KDNLNLEKKKLSYLSGLHLHIAELKEV 1160
QY 1017 LKECNKNKILFYSGANIHQFNRSRFPEDKAKESKLLWIDKEDLEKVLTKD-----NOY 1071
Db 1161 IK---NKN---YTGNSPSENNT---DVNNALESYKKELPBGTDVATVVSSESGDTLEQS 1210
QY 1072 QTERVEYRAIAGNTNERTMISTLSPGNCYCVNSIYINDSKTPISLYKKLFIISFNSFVF 1131
Db 1211 QPKKPASTHVGAESNTIT-----TSQNVDDDEVDDV-----IIVPIFGESEE 1251
QY 1132 DP-----LLRFRFVDSNVLK-----SCLYQCPMP---QPEKEILSNPLYNLAK 1172
Db 1252 DYDDJGQVVTGAVTSPVIDNILSKIEYEVLYLPLAGVYKSLKKQLENNVMTFNVV 1311
QY 1173 NTSLLIAKNDPBNFKYLLYLEYFKF-DKEKVNKILKLDKEDEF-FKEKENENNFIASLY 1230
Db 1312 KOLNSRFNKNFNKNVLESDLIPYKDLTSSNVVVK---DPYKFLNKEKDKFL--SSY 1365
QY 1231 SLAKEDFIT-----LLGDFKALKNKKGE-----DYISSLIKGVNDYL--LNN 1271
Db 1366 NYIKDSIDTDINFANDVLGYKILSEKYSKSLDSIKKYINDKQGENEKYLPFLNN 1420

RESULT 6
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
; APPLICANT: Long, David M.
; APPLICANT: Metz, Anneke M.
; APPLICANT: Love, Ruschelle A.
; TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
; FILE REFERENCE: 47714-5009-US
; CURRENT APPLICATION NUMBER: US/09/417,485D
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2184
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (330)..(335)
; OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
; OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6

Query Match 4.8%; Score 315; DB 2; Length 2184;
Best Local Similarity 20.0%; Pred. No. 1.6e-11;
Matches 323; Conservative 251; Mismatches 528; Indels 510; Gaps 82;

QY 3 TNDIVKTNPNISLYKQSKDFIKGNINIKLDFILIRKLNKPSIDNTEANIESLLKY 62
DB 233 TNNIIINNNN-----NNNNNNNNNNVHGFCHI--NNLFS-SNEPPSSNISSC--- 278

QY 63 IPELANSYVEQKAGQIEGV-----ESRVDILLPEND-----KDKAS----- 99
DB 279 -----TNYTEKNDKLTTHIRETSLITENSSKKOLLPEIDFSDRKEKSSVGVYKXKN 334

QY 100 -----FNNKL-----KEAKK-----NNEPIEDIL-----IAEVRK--- 127
DB 335 XSNIKRPHNKINTKEKKKKWKIILNRNNIIOHNTTNKCKTFLNKGHIIPDKIENNI 394

QY 128 PTFSPAKOKLKESEDLRYLYNOYQKHGYILNGKWRLY-----DKSVLYGEKRY 180
DB 395 PLFIYDLLNVIKFS-DQTYFHNFIIDEY----KQICKQIKGCTKNDISHIITSKEN 449

QY 181 IEFNFYKIEBEKEBEQEWFLFYILIRKERYLK-----TSNVIEVEKEQIAKE-- 229
DB 450 HLFHVOKLE--NNYXHPN-----INKQLRKTILKYVYVYKPEFINNVINTFKGIYRKF 503

QY 230 --KEIIQKTLKEILYERPDSDIVFKIAKNIYDKEF-----KLSGKEITQHILASILE--E 280
DB 504 PRHILNKIHK-----IFKIRLQIYIKYRIINIRMRKFIKQKVYTFPKNYD 552

QY 281 SIIFILRIF--FIAYIENDIPFKILOEN---KLYRSSI-SFRYPFYDE----- 323
DB 553 FLGSPKTYKIINFVYITKCIPIKLGSKHFKIPLKNVKKFLLFNYKESPSLNQVMK 612

QY 324 NTKKLEYKLIIFNLKDGSDAIKPPVNGGLFS---EDKVYLNNEGLLSISEEBI 380
DB 613 NIKYKNIQKIKSKYN-----IKNRILLKNIIFDNNYENKILHRNNKEI--ITIND- 661

QY 381 LVKMLPPEEKNIKDEKPVK--YSRIDPKSGFELYETLLEYDLR-----IADTTVH 428
DB 662 NIKIYKNDNLNNSPKIKTTLNKLRRKYFNKINIAIQKHLNRLIYFLPNYFIM 721

QY 429 RIIBDGYLLIRTEBELE-----NKV-----NKIATYLGNIYILTSRSLDRKSGAYYTPD 479
DB 722 PLIRRRFFLTKEQTLHKTIFDRIKWNHFTKISNFCLYHQIFRNKXKIKRNEPKM--D 778

QY 480 DLTDPMWISSIEBOLKTSPLDKII-----DNSCGSGHFLISCLDYLTKEVWLEDKFED 535
DB 779 YVQNMFNVYKKGKIKITNKYIFIKOMKKSTKNNKNNKFKKCIPIKKKKKKLNINITHNN 838

QY 536 -----VKKELDEYRVIIIESEBYDVDSI-----SKELVLKRMLL 571
DB 839 IFIKOMEKKSXTNNLINKSIDNLYK--LKEINKSVRPYIKFYYKIKKYFALKKQYI 896

QY 572 KXCIGYVDINPISVEITMLSWINTPIFGTPLSPFIEHHKAGNALLGYTKDEPFDIVKKK 631
DB 897 HMRMAKEEKSNIKLE---RAFKHFFIQAQEK--EH-----ILKYFSSHFFQNRKIN 942

QY 632 PESGFSLFKKRIKEII-----ITLEDIYOKIKGINDTTKEDIKSKYI--- 675
DB 943 YGRKFNKLIHRIKNIIIKQNSGIVKNKDKTFLHLIKNKNKNNNNKK-----KKNKNYNN 998

QY 676 -----KEYEESKDIDLRIIFSLIKLYSL-----SFDKSLNMEPSDI 712
DB 999 NINNNNNNNNNNNNNKLSNSKRYNIRNNNNKAKNNEKNIDDS-NLEKXK 1057

QY 713 ASVISLIENILGNKTSSEDEKIBKIRKLSYKFFHYGIEF----- 754
DB 1058 KIIYIKKNI-----EKNFMKLNSINHFSKLRINIPKKGRLPLINLSTL 1108

QY 755 --PDI--QEGPDVIGNPPPEKTFETEPFSKHIPNVRKLGIKEQNIILKQEILSKDNHP 810
DB 1109 NVPEIVKQRIFEIL-----KSK-KGSEBFYFHNLNLEREKKDKKNIKKKR----- 1152

QY 811 LSIYNEEKNSIIAINNIYKDPDKCFTS--GGDPNLFYFVTFNKL-IKEKGNLTLYVP 867

DB 1153 ---KYNKKNFNPVSLNNICNFSKLCGLGNRHNNSLSFKNTLTKTGBIELKJLKKWHL-- 1207

QY 868 SAINNESSIRLURKHIFARYK-----LNYIYQFENKRFKDVHSS 907

DB 1208 -KWFYKKR-MKKYIKNKLKKNKYYIYICIGDPSNVEYHNHLYFKILKNFFDNINN 1265

QY 908 KFPA-IFOLS-----NIKES-----TSSFKAQFMIOSSDNILKEITRD----- 944

DB 1266 FEFIVLFKSPRYLNKLNNSPLSYPNVVKSGFLHYIRNRELIIKSHLNDHHPLLNQ 1325

QY 945 ---LKDSKD---DAYGIELNINOI-----KKLSPIOESIIRPKONEEFTLINK 987

DB 1326 MEKTKSKSLYIFADSYKSLQVDRDIFMTIITVIRYVLYNIYVFSIKFELARKNIFYEQ 1385

QY 988 MESKPSALGEGYIDFKGLDPSIKNRKSL-----LKECNKNKLIIFYSGANIHF 1037

DB 1386 IFQE-----NOMKGYLSVRDKRVENIKKMYLNSMKKINHDEILESKNSSININ 1436

QY 1038 NSRFF-----EDKDAKESKLLWIDKEDLEKVLTKDQYQTERVFRARIASTNTERTWIST 1093

DB 1437 KKNFMCTNHEQDTEKGTQNKXHD-----IYIGPIYNSFDSF--TT 1479

QY 1094 LSPGNCYCVNSIYINDEKTPISLYKCLFIISFNSFVDFLLRRFVDSNVLSKSL---YQ 1150

DB 1480 THSSNNYKGNITHVSGD-----YKN-----DGLLHK--GNNSMNECYVKDIK 1519

QY 1151 CPMQPEEKELISNPL-----YLNKAKTSLIAKNDPENFKYLLYLE-----YFK- 1196

DB 1520 CNNN 1573

QY 1197 -----PDKEKVKNI-----LKLDKEDEF-----FKEKENENNFIIA-- 1227

DB 1574 KFLNFDLKKIISNIGYLPQGSLSNLSLYAYLDKNEEFQNLLYSEKQINNKYFLANG 1633

QY 1228 ---SLYSKAKEDFTTLGDFKALKNKKKGEDYISSLIKGYDNYLLNNKIF 1274

DB 1634 TCNYFNLSL---ILRFIDDFLFTLNKKN-----IKIFKNLLKKKIW 1674

RESULT 7
US-09-605-703B-604
; Sequence 604, Application US/09605703B
; Patent No. 6962989
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
; TITLE OF INVENTION: PROTEINS
; FILE REFERENCE: BGI-129CP
; CURRENT APPLICATION NUMBER: US/09/605,703B
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/142,764
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: 60/152,318
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 2934
; SEQ ID NO 604
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-605-703B-604

Query Match 4.7%; Score 309; DB 2; Length 622;
Best Local Similarity 19.4%; Pred. No. 7.4e-12;
Matches 125; Conservative 106; Mismatches 195; Indels 218; Gaps 20;

QY 331 YKKIITIFNLDDK-----GSDAIKFPVNGGLFSEDKVYLNNEGLLSI 374
DB 6 YDSLQLLFTLVDKHHGHTDAKAVAFDAEAGEGLHFRNLSADLFLPAATELIDRVG-LSN 64

Db 2145 KEEQHECINKFEMDFIDEVEKQKELLIKIQHLOQDCDVPSELRD-----LKLNQ 2194
Qy 963 IKKLSPIQSIIEFKDNEBFTLNKMFSEFALGEGVIDFKGLDPSIKNRKSLKCECN 1022
Db 2195 NMDLH-IEELDKDFSS-BFPSIKTEPQQV-----LSNRKEM-----2229
Qy 1023 KNLIFLYSGANIHQFNRFPEKDAKESKLLWI-----DKEDLEKVLTKDNQ--YQTERV 1076
Db 2230 -----TQFLEE-----WLNTRFDIEKLNKGIQKENDRICQVNNP 2263
Qy 1077 FYRAIASNTNERTMTSLSPGNCYCNSIYINDEKTPISLYKKLFIISFNSVFPDFLLR 1136
Db 2264 FNNRIIATNNESTEFERSA-----TISKWEQDLKSLKEKNEKL-----2303
Qy 1137 RFVDSNVLSKLYOCMPQPEEKEILSNPLYLNKNTSLIAK-NDPENFKVLLLYEYP 1195
Db 2304 -FKNYQTLKTSASGAQVNPFTQD-NKNPHVTSRA--TQLTTEKIRELENSLH-----EAK 2355
Qy 1196 KPDKEKVNKILKDKDEFFPEKENENNFIIASLYSLAKE-----DFITLLGDFFKA 1246
Db 2356 ESAMHESKLIKKQKELEV-----TNDIIAKLQAKVHESNCKLEKTKETIQVLQDKVA 2408
Qy 1247 L---KNKKGEDYISSLIK 1262
Db 2409 LGAKPYKEETEDLKMVLK 2427

RESULT 9

US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR FILING DATE: 1999-04-01
; PRIOR FILING DATE: 2000-02-01
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapSeqFormatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 4.4%; Score 285.5; DB 2; Length 2663;
Best Local Similarity 20.08; Pred. No. 1.6e-09;
Matches 292; Conservative 222; Mismatches 452; Indels 493; Gaps 72;
Qy 1 MKTNDIVKTN-----PNISLYKQLSKDFIKKENINKL-----KD 35
Db 1266 INTQLEKSHTKLQEEIPVLHEEQELLPNV---KKVSE---TQETMNELELLTEQSTTKD 1319
Qy 36 PFLIKNKLPSIDN-----STEANIESI-----LKIYFELNYSVEQQAGQIEGVSR 85
Db 1320 STTLARIEMERLRNLEKFOESQEEIKSLTKERDNLKTIKEALVYKHDQLK-----EHIRET 1375
Qy 86 VDILLFENDKDKASFNKLEAKKNPEPIEDILIIAEVKRPFTFSFDAXD-----136
Db 1376 LAKIQESQSQEQSLNKK-----EKDNETTK-----IVSEMQ-----FKPKDSALLRIEIE 1422
Qy 137 -----KLKESDOLYRYLQYQKHGYILSGKVMRYLDKSKVLYGKRYIEFNFXKIEE 190
Db 1423 MLGLSKRLQSHDWMKSVAKEDD-----LQRLQEVLSQSDQLKENIKEIVA 1470

Qy 191 KBEYKEQWFLVIYLIRKERYLKTNSVIEVEKEQIAKEKEIIQKTLKEILYERPDSDIV 250
Db 1471 KHELETEBELKVAHCCCKEQE---ETINELRVNLSKEKEIETIQKOLEAI-----NDKLO 1522
Qy 251 FKIANIYDKEPKLSGKITQHILASILEESIIFILRFFIAYIEDNDFPKILQBNKLY 310
Db 1523 NKI-QEIVEKEBOLNIQISE-----VQENV-----NEL-KQPKHEKAK 1560
Qy 311 RSSISFRPFYDENTKKLEYKKIITIFNLLDKGSDAIKPPVFGGLPSEDVKVYLNNEG 370
Db 1561 DSAL-----QSIE-SKMLELTNRLOESOEIQIMI-----KKEEMKRVQE 1600
Qy 371 LLSI-----SBIIEILVMKLFEEKNIKDEKFKVYSRLDPKPSFGELYTELTLEYDLRIAD 424
Db 1601 ALQIREDQLKENTKEIVAKM---KESQEKYQFLKTAVN-----ETQEKMEIE-----1647
Qy 425 TTVHRIEDGVYLIIRTEBELENKNVKNKIATYLNKNIYLS-----RSLDRKKS GAY 475
Db 1648 -----HLKEQFETQKLN-LENIETENIRLTQILHENLEEMRSVTKER-----1688
Qy 476 YTPDDLITDPMWISSIEEQKTKSPDLIDIKIIDNSCGSGHFLISCLDYLITKVMVELDKFED 535
Db 1689 ---DDL-----RSVEETLKVERD---QLKEN-----LRETITRDLEKQEE 1722
Qy 536 VK-----KELDBEYRVIIIE-----SEEDVDQD-SISKELVLKRLM 570
Db 1723 LKIVMHLKEHQETIDKLGIVSEKTEINSNMKOLEHNSDALKAQDALKIQBELRAHMH 1782
Qy 571 LKXCIYGV D-INPISVEITPMLSLWINTFIFGTPLSFIEHHIKAGNALLGYTKDEPFDIVK 629
Db 1783 LKEQOETIDKLGIVSEKT-----DKLSNMQKLENSNAKLEQIQEL-----1825
Qy 630 KFPESGFSLPKRIKEL---ITILEDIYQIKIGINDTK-----EDIEKSKKIYKEVESK 682
Db 1826 KANEHQLIITLKRDVNETQKVSMEQLKKQIQDQSLTSLKLIENLNLAQELHENLEEMK 1885
Qy 683 DI-----DNLRIIFSLIKL-----YSLSPDKSLNMEFSDIASVISIENILNKTSSSEDK 732
Db 1886 SVMKERDNLRRVEETIKLERDQKESLQETKARDLEIQELKLTARML-----SKEHK 1937
Qy 733 EKIEKIRKLSYYKFFHYGIBFPDIQEGFD-----IVIGNPPWE 771
Db 1938 ETVDKLE-----KISEKTIQISDIQDKLDKSKDELQKKIQELQKKELQLLRVKEDVWMS 1992
Qy 772 KTKPNETEPFSKH- -PNYRKLGIKEQNI IQEILSKDNHPLSIEYNE-----EKN 820
Db 1993 HKKINEMEQLKKQFEPNY-----LCKCEM---DNFLTKKGLHESLEIRIVAKERD 2040
Qy 821 SIITAINNIYKDFKCFCTGGDPNLPFRYFTFNLKLIKEKGN-----LTVLVP 867
Db 2041 ELNRIKESLKNERD-----QFIATIREMIARDQNHOVKPEKRLSDGQOHLME 2089
Qy 868 SAINNESSRIILRKHIFARYK-----LNYI-YQFENKCRFKDVHSSPKFAIFOLSNI 918
Db 2090 SL--REKCSRI--KELLKRYSEMDDHVECLNRLSLDLEKEIEPHRIMKKLKYVLVYVTKI 2145
Qy 919 KESTSSFKAKFMIOSSDNILKE-----ITRDLKSDKDAYKIELNINQ 962
Db 2146 KEEQHECINKFEMDFIDEVEKQKELLIKIQHLOQDCDVPSELRD-----LKLNQ 2195
Qy 963 IKKLSPIQSIIEFKDNEBFTLNKMFSEFALGEGVIDFKGLDPSIKNRKSLKCECN 1022
Db 2196 NMDLH-IEELDKDFSS-BFPSIKTEPQQV-----LSNRKEM-----2230
Qy 1023 KNLIFLYSGANIHQFNRFPEKDAKESKLLWI-----DKEDLEKVLTKDNQ--YQTERV 1076
Db 2231 -----TQFLEE-----WLNTRFDIEKLNKGIQKENDRICQVNNP 2264
Qy 1077 FYRAIASNTNERTMTSLSPGNCYCNSIYINDEKTPISLYKKLFIISFNSVFPDFLLR 1136
Db 2265 FNNRIIATNNESTEFERSA-----TISKWEQDLKSLKEKNEKL-----2304

QY 1137 RFVDSNVLSCLYQCPMPQPEKEILSNPLYNLAKNTSLIAK-NDPENFKYLLYLEYF 1195
 Db 2305 -FNYQTLKTSLSAGQVNTTQD-KNPNHVTGRA--TQUTTEKIRELENSLH-----EAK 2356
 QY 1196 KFDKEKVNKILKDKEDDFPEKEKENENNFIASLYSLAKE-----DFITLLGDFA 1246
 Db 2357 ESAMHESKIIKQKELEV-----TNDIIAKLOAKVHESKCKLEKTETIQVLQDKVA 2409
 QY 1247 L-----KNKKGEDYISSLIK 1262
 Db 2410 LGAKPYKEETEDLKMVLK 2428

RESULT 10

US-08-392-625-20
 ; Sequence 20, Application US/08392625
 ; Patent No. 5837485
 ; GENERAL INFORMATION:
 ; APPLICANT: Entian, Karl-Dieter
 ; APPLICANT: G tz, Friedrich
 ; APPLICANT: Schnell, No. 5837485bert
 ; APPLICANT: Augustin, Johannes
 ; APPLICANT: Engelke, Germar
 ; APPLICANT: Rosenstein, Ralf
 ; APPLICANT: Kaletta, Cortina
 ; APPLICANT: Klein, Cora
 ; APPLICANT: Wieland, Bernd
 ; APPLICANT: Kupke, Thomas
 ; APPLICANT: Jung, G nther
 ; APPLICANT: Kellner, Roland
 ; TITLE OF INVENTION: Biosynthetic Process For The Preparation
 ; TITLE OF INVENTION: Of Chemical Compounds
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/392,625
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/876,791
 ; FILING DATE: 30-APR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 0652.0980002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 990 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-392-625-20

Query Match 4.3%; Score 280; DB 1; Length 990;

Best Local Similarity 20.0%; Pred. No. 1e-09;

Matches 243; Conservative 186; Mismatches 394; Indels 390; Gaps 64;

QY 6 IVKTNPNISLYKQSKDFKKNINKLKDFLLKKNLKF-----SID- 48

Db 16 MVRTPIFSIELYQ-----FLKSDNI-----DYDLILQNDIFPKESIMTTTNYLQSIGKIDW 67

QY 49 --DNSTEANI--BSLLKYIIEELNYSVEQOQAGQIGVESRVDILLPNDKDKASFNKLK 105
 Db 68 EKDNKTRNVKESLLKYL---IRMSTRSTPYGMLSGV-----ALG 104
 QY 106 EAKKNNEPIPIEDILLIAEVRKPTPSFADKDKLKESEDOLYRYLNOYQKHGILSNGKW 165
 Db 105 ERSENN--IKID-----SSFHKD-VKIDGOWLYKLVHLYESDY----- 142
 QY 166 RLYDKSVLYGKRYIEFNFXXIEEKEBYKEQSWFVLYLIRKERYLKTNSVIEVEKEQ 225
 Db 143 TYKDSFVWQQNYIYNNRLYLDNNSITEN-----KENDVLSVKYNS 186
 QY 226 IAKEKEIIQTKLEILYERPDSDIVFKIKNYIDYKFKLGSKEITHOILASILEESIIFI 285
 Db 187 IL-----VFTHENSKNITVEELVQLISSKYSIENKEEVFVQOELINKELIIF- 234
 QY 286 LRIFFIAYIEDNDIFKKILOENKLYRSSISFRVFFYDENTKKLEYKIIITINLADKGS 345
 Db 235 -----SDLRPTLENKNPL-----DYINSLAPKNSL-----VGTLINISNEIT 272
 QY 346 DAIKPFVNGGLFSEDKVYLNNEGILLSISEIEELVQMLFFEEKNIKDEKFKVYKR--- 402
 Db 273 KYSKMPLGKG-----EYKYLDIVNLMS-----QLFVSKNYLQIDTYIDYSRNL 316
 QY 403 -----LDPKSPGELYETLLEVDLRIADTTVHRIIEDGVYLRTEEELE 445
 Db 317 KQSLADNISEAAAYILWLLSPHEFGT--KTIRNY-----HEFFMDKYGF-----EQLV 361
 QY 446 NKK-----VNKIATYLGNIYLTSLD-----RKKSGAYTTPDLDTDF---MWI 487
 Db 362 NLKQLSLDINGFG-YPKKDSYSFSNNIAFLKELYLAIONNHSIETENDVKNLEKNTV 420
 QY 488 SSIEEQLKT-----KSLDIKIIDNSCGS-----GHFLISCLDYLTE 524
 Db 421 SKINAPVSTEIYSEIVFGNSIKGYEDFAVISPLGSGFNAGATGRTGNF----- 470
 QY 525 KVVYELDKFEDVKELDEYRVIIIESEBYDVQDSISKELVLK-----RMLLKXCIYV--- 577
 Db 471 ----NIKKKNQLQKEIVHHYNNYMNEN---DLEISQLNEAPLNSRVNINNNRYNTCL 523
 QY 578 -----VDINPISVEITMLSIWINTFIFGTPLSFIEHHIKAGNALLGYTKDEFDVK 629
 Db 524 NLNLPKSDIDINDIFIGATFNKLYL-----YSEKH-----DSRIVFVNSMFN--- 566
 QY 630 KKFESGSLFKKRIKEIITILEDI-YOKIKINGINDTTKEDIEK-----SKKIYKEY----- 678
 Db 567 --YFSGSELYK-----FLREISPEKTKFTQITEEGIDSLPFCPRIYKNILKPAT 616
 QY 679 -----EESKDIDNLRIFSLIKYLSLSPKSNMFSDIASVLSLIENILGNKTSSEDK 733
 Db 617 WKINSEMFSETENLNRFPATIRKWHI--PKDVIIAFGD-----NRLLLNLDKHLILKK 670
 QY 734 KTEK---IRKLSYKFFHYGIEFPDIQSGFDIVIGNPPWEKTKFNETEFF-----SKHI 785
 Db 671 ELKGGHRIRESF-----INESNNRMLEIV--TPLYKTSLSKEQSFIIIPKNRKHG 721
 QY 786 PNVR-----KLGIK---EQNIIOKEILSKDNHPLSTSEYNEKNSIATINN-IKVPDFKCF 836
 Db 722 NNLKDWFSIHLSPKTYQDNFIQDYL-----PFITE-----LKVNFINKFYIYKF 768
 QY 837 TSGGDPNLPFRYFVTNFKLIIKEKGNLYLVPASAINWNESSRIIRKHIFARYKLYN---IY 893
 Db 769 KEDED-----FI--KLRLREDEYDQIY-SFIKNWKDYCLLNSLYDYSIVDYVPEV 819
 QY 894 QPENKRFKDVHSSPKF-AIFOLSNIKESTSFKAFKMIQSSNII---KEITRDKD--- 947
 Db 820 RYGGPHVIEDIENFFMYDSLDSIIIOSEFKIPKEFIVAISIDFLDLLEYLINSKSEKEI 879
 QY 948 ---SKDDAYKGLN--INQIKKL-SPIQESII---EFDNEEFTLINMFKSFKSALGEG 998
 Db 880 LINNAEDLYRSNDIREYKNLLAK/TNPKNDYELKKEFPNLHEF-----LFNKISILE-- 932

QY 999 YIDFKGLDPSIKRKSLLKCNKNLIFLYSGANIHQFNRFEDDKAKESKLLWIDK 1058
 Db 933 --NLKXTLQSLYTSRSRI-----IGSFHMRNCRIFGINPEKEFVL----- 973
 QY 1059 EDLEKVLTKDQY 1071
 Db 974 -SIFNEITTKY 985

RESULT 11

US-08-466-961A-20
 ; Sequence 20, Application US/08466961A
 ; Patent No. 5843709
 ; GENERAL INFORMATION:
 ; APPLICANT: Entian, Karl-Dieter
 ; APPLICANT: G tz, Friedrich
 ; APPLICANT: Schnell, No. 5843709bert
 ; APPLICANT: Augustin, Johannes
 ; APPLICANT: Engelke, Germar
 ; APPLICANT: Rosenstein, Ralf
 ; APPLICANT: Kaletta, Cortina
 ; APPLICANT: Klein, Cora
 ; APPLICANT: Wieland, Bernd
 ; APPLICANT: Kupke, Thomas
 ; APPLICANT: Jung, G nther
 ; APPLICANT: Kellner, Roland
 ; TITLE OF INVENTION: Biosynthetic Process for the Preparation of
 ; TITLE OF INVENTION: Chemical Compounds
 ; NUMBER OF SEQUENCES: 42
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Sterne, Kessler, Goldstein & Fox
 ; STREET: 1100 New York Avenue, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/466,961A
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/392,625
 ; FILING DATE: 22-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/876,791
 ; FILING DATE: 30-APR-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/784,234
 ; FILING DATE: 31-OCT-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Esmond, Robert W.
 ; REGISTRATION NUMBER: 32,893
 ; REFERENCE/DOCKET NUMBER: 0652.0980004
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 990 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; US-08-466-961A-20

Query Match 4.3%; Score 280; DB 1; Length 990;
 Best Local Similarity 20.0%; Pred. No. 1e-09;
 Matches 243; Conservative 186; Mismatches 394; Indels 390; Gaps 64;
 QY 6 IVKTNPNISLYKLSKDFIKKENINKLKOFFILIKNKL-----SID- 48

Db 16 MVRTEFISIELYNQ-----FLKSDNI-----DYDLILQNDIFKESIMTTTYNLYQSIGKIDW 67
 QY 49 --DNSTANI--ESLLKYIFEEELNYVEQKQKAGIEGVESRDVILLFENDKDKASFNNK 105
 Db 68 EKONKKTNRVRESLLKYL---IRMSTRSTPYGMLSGV-----ALG 104
 QY 106 EAKKNEPIPIEDILIIAEVKKRPTFSFADKDKLKESEDOLYRYLNQYQKHGILSNGKVM 165
 Db 105 EFSENN--IKID-----SSPHKKD--VKIDGQWLYKLVHVESDY----- 142
 QY 166 RLYDKSKVLYGKRYIEFNFXKIEBEKYEQEWFLFYILIRKERYLTSTNVIEWEKEQ 225
 Db 143 TYKDSFVIWNOQNTYNNRLYLDNNSITEN-----RRNDVLSVKYNS 186
 QY 226 IAKEKEIIQKTLKEILYERPDOSIVFKIAKNYIDKEFKLSGKEITQHILASILEESIIFI 285
 Db 187 IL-----VFIHNSKKNITVEELVQIISKYSIENKEEVKFVQOELINKELIIF-- 234
 QY 286 LRIFFTAVIEDNDIFPKILQENKLYRSSISPRYFFYDENTKKKLEYKIIITIFNLDKGS 345
 Db 235 -----SDLRPTLENKNPL-----DYIINSLNPKNSL-----VGTLINISNEIT 272
 QY 346 DAIKFPVFNGLFSEDVKVYLNNEGLLSISEIEILVKMLFFEEKNIKDEKPFVKYSR--- 402
 Db 273 KYSKMPLGKG-----EYKYLDIVNLS-----QLFVSKNYLQIDTYIDYSRNL 316
 QY 403 -----LDPKSFGELYETLLBYDLRIADTTVHRRIEDGVYLIIRTEEELE 445
 Db 317 KQSLADNISEAAAYILWLSPHEFGT--KTIRNY-----HEFPMDKYGF-----EQLV 361
 QY 446 NKK-----VNKIATYKGNILYTSRSLD-----RKKGAYVTPDDLTDF---MWI 487
 Db 362 NLKQLSLDINGFG--YPKDSYSFSNNIAFLKKEYLLAIONNSHIEITENDVNLEKNIV 420
 QY 488 SSIEEQLKT-----KSLPDIKIIDNSCGS-----GHPLISCLDYLTE 524
 Db 421 SKINAPVSTEIYSEIYFGNSIKGYEDFAVISPLGSGFNAGATGPRGTGNF----- 470
 QY 525 KVMYELDKPEDVKELDEEYRVIIIESEBYDVQDSISKELVLK-----RMLLKCIYG--- 577
 Db 471 ----NIKKQNQLQKEIVHHNNYNNMEN---DLEISOLNEAPLNSRVNLIANNRNIYNTCL 523
 QY 578 -----VDINPISVEITMLSLWINTFIFGTPLSFIEHRIKAGNALLGYTDEFFDIVK 629
 Db 524 NLNLPKSDIDINDIFIGATFNKLYL-----YSEKH-----DSRIYFVSNMFEN--- 566
 QY 630 KXFESGFSLPKKRIKEIITILEDI--YQKIKGINDTTKEDIK-----SKKIYKEY----- 678
 Db 567 --YFSGSELYK-----FLREISPEKTKFIQITEEGIDSLPFCPRIIYKNIILKPAT 616
 QY 679 -----ESKDDIDLRIIFSLIKLYSLSPKSLNMEFSDIASVISLIENTLGNKTSSEDE 733
 Db 617 WKINSEMFSETENLNRNFATIRKWHI--PKDVIIFAFD-----NRLNLNLDKHLILKK 670
 QY 734 KIEK--IRKLSYKFFHYGIEFPDIQGFQDIVIGNPPWEKTKFNETEFF-----SKHI 785
 Db 671 ELKHGRIRILESP-----INESNNRMLEIV--TPLYKKTSLKEQSIIIPKRNKHF 721
 QY 786 PNYR-----KLGIK---EQNIKQEILSKDNHPLSLEYNEEKNSIIAINN--IYKFPKCF 836
 Db 722 NNLKDWFSIHLSIPKTYQDNFIQDYLL-----PFITE-----LKVANNFINKFFYIKF 768
 QY 837 TSGGDPNLFRYPVTFNLKLIKKEGNTYLVPSAINWESSRILRKHFARYKLVN---IY 893
 Db 769 KEDED-----FI--KLRLLEDEDSQIY--SPKKNWKDYCLNSELSDYSDIYDVEVY 819
 QY 894 OPENKKRKFQDVHSSKFF--AIFQLSNIKESTSSPKAKFMIOSSDNIL---KEITRDLKD-- 947
 Db 820 RYCGPHVIEDIENFFMYDSDLSINIIQSEFKIPKEFIVAISIDFLDLYLEINKSEKEI 879
 QY 948 ---SKDDAYKGIELN--INQIKKL-SPIOESI---EFKDNESFTLKNKMFSPKSALEG 998

Db 880 LINNAEDLYSNDIREYKLLAKLTNPKNDEYELKKEFPNLFHFF-----LFNKISILE-- 932

Qy 999 YIDPKGLDPSIKRKSLLKECNKNKLI FLYSGANIHFQNSRFFEDKAKESKLLWIDK 1058

Db 933 --NLKTKTQSLATSRRI-----IGSFIMRCNRIFGINPEKEKFUL----- 973

Qy 1059 EDLEKVLTKDNQY 1071

Db 974 -SIFNEITTKYKY 985

RESULT 12

US-08-645-193B-15

Sequence 15, Application US/08645193B

Patent No. 5962253

GENERAL INFORMATION:

APPLICANT: Kupke, Thomas

APPLICANT: Gotz, Friedrich

APPLICANT: Kempter, Christoph

APPLICANT: Jung, Gunther

TITLE OF INVENTION: Oxidative Decarboxylation of Peptides

TITLE OF INVENTION: Catalyzed by Flavoprotein EpiD

NUMBER OF SEQUENCES: 70

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

STREET: 1100 New York Avenue, Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/645,193B

FILING DATE: 13-MAY-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.

REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 0652.1540000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 990 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: not relevant

MOLECULE TYPE: protein

US-08-645-193B-15

Query Match 4.3%; Score 278; DB 1; Length 990;

Best Local Similarity 20.1%; Pred. No. 1.3e-09;

Matches 243; Conservative 183; Mismatches 402; Indels 380; Gaps 63;

Qy 6 IVKTNPNISLYQLSKDFTKGNINIKLDPFILRNKLP-----SID- 48

Db 16 MVRTPIFSIYLNQ-----FLKSDNI-----DYDLILQNDIFKESIMTTTNYLQSIGKIDW 67

Qy 49 --DNSTEANT-ESLLKYIFELNYSVEQKAGQIEGVESRVDILLFENDKDKASFNK 105

Db 68 EKONKTRNKESLLKYL-----IRMSTRSTPYGMLSGV-----ALG 104

Qy 106 EAKKNPEPIEDILIIAIEKRPFTSPDAKDKLKESEDQLYRYLNQYQKHGILSNK 165

Db 105 EFSENN- IKKD-----SSPHKD-VKIDQWLYKLHVHLESY----- 142

Qy 166 RLYDKSKVLYGKRYIEFNFKIEEYKEQBWFLFIYLIRKERYLKTNSVIEVEKEQ 225

RESULT 13

US-09-957-005-9

Db 143 TTYKDSFVWQNYIYNNRLYLDNNSSITEN-----KENDVLVSKYNS 186

Qy 226 IAKEKEIIOKTLKEILYERPDPSIVFKIAKNIYDKFKLSGKEITOHILASILEESIIFI 285

Db 187 IL-----VFTHENSKNITVEELVQLISSKYSTENKEEVVQELINKKEIIF- 234

Qy 286 LRIFFIAYIEDNDIFKKILQENKLYRSSISFRFFYDENTKKKLEYKKIITINLLDKGS 345

Db 235 -----SDLRPTLENKNPL-----DYIINSLNPKNSL-----VGLTINISNEIT 272

Qy 346 DAIKFPVFNGLFSEDEKVKYLNNEGLLSISEBEILVKMLFFBEKNIKDEKPKVYSR-- 402

Db 273 KYSKMPLGKG-----EYKYLDIVNLMS-----QLFVSKNYLQIDTYIDYSENEL 316

Qy 403 -----LDPKSGELYETLLEVDLRIADTTVHRIIEDGVYLIRTEEELE 445

Db 317 KQSLADNISGAAYILMLSPNHFGT--KTIRNY-----HEFFMDKYGF-----EQLV 361

Qy 446 NKK-----VNKIATYLGKNIYLTSLD-----RKSGAYTTPDLDTF--MWI 487

Db 362 NLKQLLSDINGFG-YPKODSYGSFNNAFLKEKYLLAIQNNSHIEITENDVKNLEKNNTV 420

Qy 488 SSIIEOLKT-----KSPLDIKIIDNSCGS-----GHFLISCLDYLTEKVMY 528

Db 421 SKINAPVSTIEYSEIYFGNSIKGYEFAVISPILGSGFNAGATGRTGN-----F 470

Qy 529 ELDPEDVKKELDEEYRVIIERSEYDVQDSISKELVLK-----RMLLKCIYG----- 577

Db 471 NIKKKNLQKEIVHHYNNYMNEN---DLEISQLNEAPLNSRVNVLNNRIYNTCLNML 527

Qy 578 -----VDINPISVEITMLSWINTFIRGTPLSFIEHHIKAGNALLGTYKDEFFDIVKKKE 633

Db 528 PKSDIDINDIFGATFNKLYL-----YSEKH-----DSRIVFVSMSFN-----YE 568

Qy 634 SGFSLPKKRIKEIITILEDI-YOKIKGINDTTKEDIK-----SKKIYKEY----- 678

Db 569 FGSELYK-----FLREISFEKTFIQPIIEEGIDSLPFCPRIYIKNILLPATWKN 620

Qy 679 -BESKDIDNLRIFSLIKLYSLFDSKLNMEFSDIASVISLIENILGNKTSSEDEKIEK 737

Db 621 SEMFSETENLNRFATIR-EKWHIPKDVIIAFGD---NRLNLLNLDKHLIILKELKK 675

Qy 738 ---IRKLSYKFFHYGIEFPDIOEGFDIVIGNPPWEKTKFNTEFP-----SKHIPNYR 789

Db 676 HGRIRILESP-----INESNNERMLEIV--TPLYKTSLSKQSFIIKNNRKHFNK 726

Qy 790 -----KLGIK---EQNIKQELISKDNHPLSIEYNEEKNSIIAINN-IYKFDKCFCTSGG 840

Db 727 DWFSIHLSIPKTYQDNFIQDYLL-----PFITE-----LKVNNPINKFFYIKFEDE 773

Qy 841 DPNLFRYFVTNKLKIKKGNLTLYVPSAIWNSSSRILRKHFARYKLYN---IYQFEN 897

Db 774 D-----FI--KLRLREDEYQIY-SPIKWKDYCLLNSELYDYSIVDYVEVRYGG 824

Qy 898 KRFKDVHSSFKPAIFOLSNIKESTSSFKAKFIQSSDNIL---KEITRDLK-----SK 949

Db 825 PHVIEDIENFFWYDLSLSINIIQSEFKIPKEFVAISIDFLDLYLEINKSEKEILINNA 884

Qy 950 DDAKGIELN--INQIKKL-SPIQESII---EFKDNBEETLINKMSEKFSALGEGYIDFK 1003

Db 885 EDLYRSNDIREYKNLAKLTNPKNDEYELKKEFPNLFHFF-----LFNKISILE-- 935

Qy 1004 KGLDPSIKRKSLLKECNKNLIFLYSGANIHFQNSRFFEDKAKESKLLWIDKEDLEK 1063

Db 936 KTLQKSLYTSRRI-----IGSFIMRCNRIFGINPEKEKFUL-----SIFN 977

Qy 1064 VLTKNQY 1071

Db 978 EITKTKKY 985

Query Match 4.1%; Score 269; DB 1; Length 2710;
Best Local Similarity 19.4%; Pred. No. 1.9e-08;
Matches 304; Conservative 257; Mismatches 499; Indels 506; Gaps 82;

QY 1 MKTNDIVKNNPNISLYKQLSKDFIKENINKLKDFILLIKNKLFGIDDNSTEANTESLL 60
DB 492 LKASDLIEKFP-----ENLSQLEQEI---NSLWFDQASAKYQPE--- 531
QY 61 KYIFEELNVSBOQAGQIEGUESRVDILLFEND--KDKASFNNKLKEAKKNPEPIED 118
DB 532 KYVRDYTGGSLSBD-----NGVDFNKNTALDKNYLLNNKIPSNVVEAGSKY- 579
QY 119 ILIIAERKPTESFADKDLKESEDQ---LYRYLNOYQKHGYILSGKVRWLYDKSKVL 174
DB 580 VHILOQDDISYEATCNLFSPKNKNSIIQRMNSESASYPFLSDG----- 627
QY 175 YGEKRYIBFNFKIEKEBYKEQEWFLFYIIRKERYLKTSNVIEVEKEQIAKEBIIQ 234
DB 628 ---ESILELNKYRIPERLKNKEK---VKYTFIGHKDEFNTESEFARLSVDSLSNEISSFL 681
QY 235 KTLKEILYERPDOSIVFKIAKNIYKBFKLS-----QKEITQHILASILEE----- 280
DB 682 DTIK--LDISPKNVENVLLGCNMFSDFNVEETYPGK-----LLLSIMDKITSTLPDVNK 734
QY 281 -SIIFILRIFFAYIEDNIFPKILOENKLYRSSISFRYFFYDENTKKLEYKIIITFN 339
DB 735 NSIT-----IGANQVEYRINSEGRKELAHSGKWINKEEATMSLSSKEYI--FFD 783
QY 340 LLDKGSDAIKFPVFNGLFSEDKVYLNNEGILSIEIEILVKMLFPEEKIKDKBFVK 399
DB 784 SID-----NKLKAKSNIPGLASISED-----IKTLLDASVPTKPFIL 823
QY 400 YS-RLD-PKSPGB-LYETLLEVDLRIADTVVHRIIEDGVYLIRTEBEL-ENKKNVKIATY 455
DB 824 NNLKLNIESIGDYIYKLEPKVNIHNSIDDLIDFNLENVSDLYELKKN----- 878
QY 456 LKGNILYTSRDLRCKSGAYYPPDDLTDPMVSISSIEQKTKSPDLIKIDNSCGSHFL 515
DB 879 -----NLDEK-----YLISPEDISKNNSTYSVRFINSG----- 908
QY 516 ISCLDYLTEKVMYELDKFEDVKELDEYRVIIIESEYDVQDSISIKEL-VLKRMLLKC 574
DB 909 -----BSVVE-----TEKEIFSKY-----SEHITKEISTIKNSII----- 939
QY 575 IYGVNDINPISVEITML--SLWINTFTFGTPLSFIEHHKAGNALLGYT--KDEFFDI--- 627
DB 940 ---TDVGNLLDNQILDHTSQVNTL---NAAPFIQ-----SLIDYSSNKVDNLDLSTS 986
QY 628 VK-----KKPESGF-----SLFKRKRIEIIITLEDIYQKI-----KGIN--DT 663
DB 987 VKVQLYAQLFSTGLANTIYDSIQLVNLISNAVNDTINVLTPIITEGPIVSTILDGINLGA 1046
QY 664 TKEDIEKSKIKYKEYESKDIDNRIIFSLIKLYLSFDKSLNMEFS-----DIASVISLI 719
DB 1047 IKELDEHDFLLKKELEAK--VGVLAINMSLSIAATVASIVGIGAEVTIFLLPPIAGISAGI 1105
QY 720 ENILGNKTSSEK--BKIEKIRKLSYYFFHYGIEFFDQIEGF-DIVIGNPPEWKTKE- 775
DB 1106 PSLVNNELIHDKATSVNVFNHLSSEKYGKPLKTEDDKILVPIDDLVISEDFNNNSIK 1165
QY 776 -----NETEFF-----SKHPN---YRKLGIKEQNI-IKQETLSKD 807
DB 1166 LGTCNILAMEGGSGHTVTGNDHFFSSPSISSHIPSLSIYSAIGIBTENLDFSKIMLPL 1225
QY 808 NHPLSTIEY-----NEEKNSIITAINIY--KFDPKFTSGGDPNLPFRYFT-- 850
DB 1226 NAPSFRVFWETGAVPGLRSLENDGTLLDSIRDLYPKFYWRFYA-----FFDYAITTL 1279
QY 851 -----FNLKLIKEKGNLTVLPSPAIVNESSRILKHIPARYKLVNIYQFENKKEFKDV 904
DB 1280 KPVEDTNIKIKDKDTRNFIPTITNE-----IRNKLGSYDFGAG-GTYSLL 1327
QY 905 HSSFKFAIFQLSNIKESTSSFAKFMIOQSDNLIKETITRD---LKDSK--DDAYKGIELN 959

DB 1328 LSSYPIS-----TNINLS-----KDDLWIFNIDNREVEISIENGITKKGKLIKDVLSKIDIN 1379
QY 960 INQIKKLSPIQBSIIEFK---DNE-----EFTLNKMFKFSKALGEG 998
DB 1380 KKKL---IIGNQITDFSGDINDKORYIFLTCELDKISLIIEINLVAKSYSLLSGDKN 1435
QY 999 YDFKFGGLDPSIKNRKSLKECN-----NKNLIFYSGANIHQFNRFEDDKAKESSKL 1053
DB 1436 YL-----ISNLSNTIEKINTILGLDSKNIAINYT-----DESNNKYP-GAISKTSQKS 1481
QY 1054 LMKIDEDLEKVL-----TKDNQVOTERVFVR-----ATASNTNERTMIS 1092
DB 1482 ITHYKDKSNILIEFYNDSTLEFNSKDFIADINVMKDDINITGKYVYVONNIDKSIDFS 1541
QY 1093 -TSPGNCYCVNSIYIN-----DEKTPISLYKKLFI--ISIFNSFVFD--- 1132
DB 1542 ISLAVSKNQVKNGLYNLSESVYSSYLDLVKNSDGHNTSNFMNLFNLDNISFWKLFGFENIN 1601
QY 1133 FLRRP-----VDSN-----VLKSLCYOC-----PMPQP 1156
DB 1602 FVIDKYFTLVGKTNLGYVEFICDNNKNIDIFYGEWKTSSSKSTIFSGNGRNVVVEIYNP 1661
QY 1157 EEKEILS-----NPLY-----LNLAKNTSL-----LIAXN- 1181
DB 1662 DTGEDISTLDSYBPLYGIDRYINKVLIADPLYTSLINI--NTNYSNEYYPEIILVLP 1719
QY 1182 -----DPENPKY-----LLYLEYFKFDKEKVNKILKL-----DKED 1212
DB 1720 NTFPHKKVNLNDSSFEYKMWSTEGSDFFILVRVLEBSNKKILQIRIKIGILSNTQSFNKS 1779
QY 1213 EFFKE-KENENNFIASLYSLAKEDFITL--LGDFKALKNKKKGEDYISLKGYNLYL 1269
DB 1780 IDFKDKIKLSLGYINMSNPKSFNSENELDRHLG-FKIIDNKTYYYDEDSKLVKGLIN--I 1836
QY 1270 NKKIFY 1275
DB 1837 NNSLFY 1842

RESULT 15
US-08-405-496A-6
Sequence 6, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992
 PRIOR APPLICATION DATA: US 07/429,791
 FILING DATE: 31-OCT-1989
 ATTORNEY/AGENT INFORMATION:
 NAME: INGOLIA, DIANE E.
 REGISTRATION NUMBER: 40,027
 REFERENCE/DOCKET NUMBER: OPD-01308
 TELEPHONE: (415) 705-8410
 TELEFAX: (415) 397-8338
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-405-496A-6

Query Match 4.1%; Score 269; DB 1; Length 2710;
 Best Local Similarity 19.4%; Pred. No. 1.9e-08;
 Matches 304; Conservative 257; Mismatches 499; Indels 506; Gaps 82;

QY 1 MKNIDIVKTNPNISLYKQLSKDFIKKENINKLKDFILIKNKLFSIDDNSTEANTIESLL 60
 DB 492 LKASDLIEFKFP-----ENNLSQTEQEI---NSLWFDQASAKYQFE--- 531
 QY 61 KYIFEELNYSVEQKAGQIBGVSRVDILLFEND--KDKASFNNKLKEAKKNNEPIPIED 118
 DB 532 KYRVDYTGGLSED-----NGVDFKNKALDKNLLNNKIPSNVVEAGSKNY----- 579
 QY 119 ILIIAEVRRPTFFDAKDKLKESEDQ---LYRYLNOYQKHGYLNGKWRLYDKSKVL 174
 DB 580 VHYIIQIQGDDISYEATCNLFSPKNPKNSIIIOQNMESAKSYFLSDG----- 627
 QY 175 YGKRYIEFNFXKIEBEKVEKEQEWFLFYILIRKERYLNTSVNIEVEKEQIAKEKIIQ 234
 DB 628 ---ESILELNKRIPELKNKEK---VKVTFIGHGKDEFNTEPFLSVDSLSNEISSFL 681
 QY 235 KTLKEILYERPDOSIVFKIAKNYDKFKLS---GKEITQHILASILER----- 280
 DB 682 DTIK--LDISPKVNEVNLGCNMFSDYFVEEYPGK-----LLLSIMDKITSTLPDVK 734
 QY 281 -SIIFILRIPIAVIENDIFKILQENKLYRSISPRYPFYDENTKKLEYKIITIFN 339
 DB 735 NSIT-----IGANQVEVRINSRGRKELLAHSGKWINKEBAINMSDLSKEYI--FFD 783
 QY 340 LLDKGSDAIKFPVNGFLPSEDVKYLNNGLLSISEIEILVKMLPFEEKNIKDEKFKV 399
 DB 784 SID-----NKLKAKSNIPGLASISED-----IKTLILDASVSPDTFKIL 823
 QY 400 YS-RLD-PKSGFE-LYETLLEYDLRIADTTVHRIEDGVYLIRTEEL-ENKVKNIATY 455
 DB 824 NNLKLNIESSIGDVIYKELEPVKNIIHNSIDDLIDFNLENVSDLEYELKLN----- 878
 QY 456 LKGNIIYTSRLDRKSGAYTTPDLDTFMVVISIEPQLTKPSPLDIKIIDNSCGSHFL 515
 DB 879 -----NLDEK-----YLIISFEDISKNNSTYSVRFINKSG----- 908
 QY 516 ISCLDYLTEKVMYELDKFEDVKKELDEYRVYIEESEYDVQDSISKEL-VLKEMLLKXC 574
 DB 909 -----ESYVE-----TEKEIFSKY-----SEHITKEISTIKNSII--- 939
 QY 575 IYGVDPINPISVEITMVL--SLWINTFIPTPLSFIEHHIKAGNALLGYT--KDEPFDI--- 627
 DB 940 ---TDVNGNLLDNTQLDHTSQVNTL---NAAFFIQ-----SLIDYSSNKVDVNDLSTS 986
 QY 628 VK-----KKFESGF-----SLFKRIKEIITILEDIYQKI-----KGIN--DT 663
 DB 987 VKVQLYAQLFSTGLNTIYDSIQLVNLISNAVNNDTINVLPITTEGIPVSTILDGINLGA 1046
 QY 664 TKEDIKSKKIYKEVEESKIDNRIIFSLIKLYSLSPDKSLNMEFS-----DIASVISLI 719

Search completed: January 24, 2006, 20:02:59
 Job time : 94.7244 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 224.44 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-16
Perfect score: 6538
Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIKGNDYLLNNKIFYHK 1277

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:*
1: /cgn2_6/prodata1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6534	99.9	1277	5	US-10-688-058-16 Sequence 16, Appl
2	872	13.3	174	5	US-10-688-058-14 Sequence 14, Appl
3	455.5	7.0	1253	4	US-10-335-977-7399 Sequence 7399, Ap
4	435	6.7	1532	3	US-09-738-626-4321 Sequence 4321, Ap
5	419.5	6.4	1279	3	US-09-882-227-388 Sequence 388, App
6	392	5.8	1009	5	US-10-800-946-4 Sequence 4, Appli
7	354	5.4	1441	5	US-10-732-923-3352 Sequence 3352, Ap
8	337	5.2	2910	5	US-10-732-923-3342 Sequence 3342, Ap
9	335.5	5.1	1939	5	US-10-732-923-3340 Sequence 3340, Ap
10	328.5	5.0	1527	5	US-10-732-923-3354 Sequence 3354, Ap
11	324	5.0	1985	4	US-10-732-923-3351 Sequence 3351, Ap
12	318	4.9	740	4	US-10-335-977-7398 Sequence 7398, Ap
13	316	4.8	1602	5	US-10-954-924-7 Sequence 7, Appli
14	316	4.8	1621	5	US-10-954-924-5 Sequence 5, Appli
15	316	4.8	1639	4	US-10-087-464-10 Sequence 10, Appl
16	316	4.8	1639	5	US-10-954-924-3 Sequence 3, Appli
17	315	4.8	2184	4	US-10-304-095-6 Sequence 6, Appli
18	314	4.8	696	4	US-10-335-977-7397 Sequence 7397, Ap
19	307	4.7	1478	5	US-10-732-923-3353 Sequence 3353, Ap
20	301.5	4.6	1163	4	US-10-282-122A-51864 Sequence 51864, A
21	301.5	4.6	1163	5	US-10-732-923-3296 Sequence 3296, Ap
22	295.5	4.5	1875	4	US-10-369-493-22285 Sequence 22285, A
23	295.5	4.5	1875	5	US-10-732-923-3334 Sequence 3334, Ap
24	295.5	4.5	1875	5	US-10-732-923-3335 Sequence 3335, Ap
25	293	4.5	1178	4	US-10-282-122A-52434 Sequence 52434, A
26	293	4.5	1196	4	US-10-282-122A-52737 Sequence 52737, A
27	293	4.5	6761	5	US-10-732-923-15035 Sequence 15035, A

ALIGNMENTS

RESULT 1

US-10-688-058-16
; Sequence 16, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCR0:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1277
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (186)
; OTHER INFORMATION: Xaa = anything
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (573)
; OTHER INFORMATION: Xaa = anything
US-10-688-058-16

Query Match 99.9%; Score 6534; DB 5; Length 1277;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKTNDIVKTNPNISLYKQLSKDFIKKENINKDKPFIILIKNKLPSIDNSTEANTESLL	60
Db	1	MKTNDIVKTNPNISLYKQLSKDFIKKENINKDKPFIILIKNKLPSIDNSTEANTESLL	60
QY	61	KVIFELNYSVQKQAGQIEGVESRVDILLFENDKDKASFNNKLKEAKKNNEPIEDIL	120
Db	61	KVIFELNYSVQKQAGQIEGVESRVDILLFENDKDKASFNNKLKEAKKNNEPIEDIL	120
QY	121	IIAEVKRPTFSFDKDKLKESEDQLYRYLNOYQKHGILSNQKWLVDKSKVLYGEKRY	180
Db	121	IIAEVKRPTFSFDKDKLKESEDQLYRYLNOYQKHGILSNQKWLVDKSKVLYGEKRY	180
QY	181	IEFNFKIEEKEEYKEQEWFLFIYLIRKERVYKTSNVIEVEKEQIAKKEIIQKTLKEI	240
Db	181	IEFNFKIEEKEEYKEQEWFLFIYLIRKERVYKTSNVIEVEKEQIAKKEIIQKTLKEI	240

QY 241 LYERPDSIVFKIAKNYDKEFKLSGKEITQHTLASILEESIIFILRIFFIAYIENDIF 300
DB 241 LYERPDSIVFKIAKNYDKEFKLSGKEITQHTLASILEESIIFILRIFFIAYIENDIF 300
QY 301 KKILQENKLYRSSISPRYPFYDENTKKLSYKKIITIFNLLDKGSDAIFPPVNGGLFSE 360
DB 301 KKILQENKLYRSSISPRYPFYDENTKKLSYKKIITIFNLLDKGSDAIFPPVNGGLFSE 360
QY 361 DKVKYLNNEGLLSISIEEILVKMLPFEEKNIKDEKFKVYSRLDPKPSFGELYETLLEYDL 420
DB 361 DKVKYLNNEGLLSISIEEILVKMLPFEEKNIKDEKFKVYSRLDPKPSFGELYETLLEYDL 420
QY 421 RIADTTVTHRIIEOGVVLIRTEEBLENKVNKIATYLGKNIYLTLSRLDRKKS GAYTTPD 480
DB 421 RIADTTVTHRIIEOGVVLIRTEEBLENKVNKIATYLGKNIYLTLSRLDRKKS GAYTTPD 480
QY 481 LTDPMVISSIEQLKTKSPLDIKIIDNSCGSGHFLISCLDYLTEKVMYELDKFEDVKKEL 540
DB 481 LTDPMVISSIEQLKTKSPLDIKIIDNSCGSGHFLISCLDYLTEKVMYELDKFEDVKKEL 540
QY 541 DEBYRVILIESEBYDVQDSISKELVLKRMLLKXCIIYGVNDINPISVEITMLSLWINTFI 600
DB 541 DEBYRVILIESEBYDVQDSISKELVLKRMLLKXCIIYGVNDINPISVEITMLSLWINTFI 600
QY 601 TPLSFIEHHITKAGNALLGYTKDBFFDIVKKPFESGSLFKRIKEIITILEDIYQIKGI 660
DB 601 TPLSFIEHHITKAGNALLGYTKDBFFDIVKKPFESGSLFKRIKEIITILEDIYQIKGI 660
QY 661 NDTTKEDIEKSKKIYEYESKDIDNLRIIFSLIKLYSLFSDKSLNMFSDIASVLSLIE 720
DB 661 NDTTKEDIEKSKKIYEYESKDIDNLRIIFSLIKLYSLFSDKSLNMFSDIASVLSLIE 720
QY 721 NILGNTSSBDKIEKIRKLSYKPFHYGIBFPDIQEGFDIVGNPPWEKTKFNETEF 780
DB 721 NILGNTSSBDKIEKIRKLSYKPFHYGIBFPDIQEGFDIVGNPPWEKTKFNETEF 780
QY 781 FSKHPIPNYKLGITKEQNIIOEILSKDNHPLSIYNEEKNSIITAINNIYKDFPKCFTSG 840
DB 781 FSKHPIPNYKLGITKEQNIIOEILSKDNHPLSIYNEEKNSIITAINNIYKDFPKCFTSG 840
QY 841 DPNLFYFVTFNLKLIKKEGNLTVLPSAIWNSSSRILRKHIFARYKLNYYIQENKKR 900
DB 841 DPNLFYFVTFNLKLIKKEGNLTVLPSAIWNSSSRILRKHIFARYKLNYYIQENKKR 900
QY 901 FKDVHSSFKFAIFQLSNIKESTSFKAKFMIQSSDNILKEITRDLKDSKDDAYKGIELNI 960
DB 901 FKDVHSSFKFAIFQLSNIKESTSFKAKFMIQSSDNILKEITRDLKDSKDDAYKGIELNI 960
QY 961 NQIKLSPIQESIIEPKDNEEFTLINQMFKFSALGEGYIDFKGLDPSIKNRKSLKKEC 1020
DB 961 NQIKLSPIQESIIEPKDNEEFTLINQMFKFSALGEGYIDFKGLDPSIKNRKSLKKEC 1020
QY 1021 NNKNLIFLYSGANIHQNSRFFEDKDAKSSKLMWIDKEDLEKVLTKDNQYQTERVFPYRA 1080
DB 1021 NNKNLIFLYSGANIHQNSRFFEDKDAKSSKLMWIDKEDLEKVLTKDNQYQTERVFPYRA 1080
QY 1081 IASNTNERTWISTLSPGNCVNSIYINDEKTPISLYKKLFIISFNSFVDFLLRRFVD 1140
DB 1081 IASNTNERTWISTLSPGNCVNSIYINDEKTPISLYKKLFIISFNSFVDFLLRRFVD 1140
QY 1141 SNVLKSLCYOCMPQPEKEITLSNPLYNLAKNTSLIAIKNNDPENPKYLLYLEYFKPDKE 1200
DB 1141 SNVLKSLCYOCMPQPEKEITLSNPLYNLAKNTSLIAIKNNDPENPKYLLYLEYFKPDKE 1200
QY 1201 KVKKILKLDKDEDFEKEKENENNFIASLYSLAKEDFITLIGDPKALKNKKKGEDYISSL 1260
DB 1201 KVKKILKLDKDEDFEKEKENENNFIASLYSLAKEDFITLIGDPKALKNKKKGEDYISSL 1260
QY 1261 IKGYDNYLLNNKIFYHK 1277
DB 1261 IKGYDNYLLNNKIFYHK 1277

RESULT 2
US-10-688-058-14
; Sequence 14, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-14
Query Match 13.3%; Score 872; DB 5; Length 174;
Best Local Similarity 99.4%; Pred. No. 6e-42;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 268 EYQHILASILEESIIFILRIFFIAYIENDIFKKILOENKLYRSSISPRYPFYDENTKK 327
DB 1 QYQHILASILEESIIFILRIFFIAYIENDIFKKILOENKLYRSSISPRYPFYDENTKK 60
QY 328 KLEYKKIITIFNLLDKGSDAIFPPVNGGLFSEDVKYLNNEGLLSISIEEILVKMLPF 387
DB 61 KLEYKKIITIFNLLDKGSDAIFPPVNGGLFSEDVKYLNNEGLLSISIEEILVKMLPF 120
QY 388 EKNINDEKFKVYSRLDPKPSFGELYETLLEYDLRIADTTVTHRIIEOGVVLIRTE 441
DB 121 EKNINDEKFKVYSRLDPKPSFGELYETLLEYDLRIADTTVTHRIIEOGVVLIRTE 174
RESULT 3
US-10-335-977-7399
; Sequence 7399, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 7399:
SEQUENCE CHARACTERISTICS:
LENGTH: 1253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...1253
SEQUENCE DESCRIPTION: SEQ ID NO: 7399:
US-10-335-977-7399

Query Match 7.0%; Score 455.5; DB 4; Length 1253;
Best Local Similarity 20.7%; Pred. No. 3.4e-17;
Matches 302; Conservative 202; Mismatches 499; Indels 453; Gaps 61;

QY 22 KQPIKEN-----INKLKDFFILIKNKLPSIDDNSTEANIESLLK--YIFEELNYS 70
DB 10 KVFIKYNQEPKKTETIENFEKINSLEENVPQDDDEEFQKNEINKFLKNAYGYDNTY- 68
QY 71 VEQOKAGQIEGVESVDIILLFENDKDKASFNNKLKKAENNEPIP-----IEDI- 119
DB 69 ---EKVDYIYVDKEVRVLI-----EVKALNNKTEPFPKNRENPLSKAFQWLVLYFLEEIE 120
QY 120 -----LIIAEVVRPFPSPDAKDKLKESED-OLYRVLYNOYKHGILSNGKWRLYD 169
DB 121 KEKNSLKHITIICNA-HEFPLFCKDLLSLKEDKRIKKFYKNYAKKEGT-----D 169
QY 170 KSKVLY-----GKRYIEFNFAXKEEKEEYKEQWFLVIYLIRKERYLKTS 216
DB 170 SSKPKPKYKLEQFLOQSDFOGELRYTHFNL-----SDDFKS---LPLIYQVLSQEVLLKQK 221
QY 217 NVIEVEKEQIAKEKEIIQKTLKEILYERPD--SIYFKIA--KNYIDKEFKLSGKEITH 272
DB 222 RTLDAN-----TLNKQFYELLYTLGLEBQNDKGLIKILPRTKNSLSGALKKRYKNLDD 277
QY 273 ILASILEESIFILRIFFIAYIEDNDI FKKILOENKLYRSSISFRYF-----FYDENTK 326
DB 278 EVWALL--IANNRILFLRLLESLLISFEHFEKERSFLTTGNFKDFNMLNTLFFEVLLAK 334
QY 327 KCLEYKIIITIFNLLDKGSDAIFPVPNGGLFSE-----DKVKYLNNEGLLSISEIEE 379
DB 335 KNSRLPEIKENKILG-----KIPYLNSSLFPDKTPLELKKGHEIKLLDNKKL-----E 381
QY 380 ILVKMLFFBEKNIKDKFKVYSRLDPKSPGELYETLLEYDLRIAD-----TTVHRITEDGV 435
DB 382 IYKNSVLKKNHYQDK-----ASLPLLEFFFKFLHYKFTTTPKOIKDNT 427
QY 436 YLIRTEEELENKVNKIATYKGNILYLSRDKKSGAYTTPDDLTDENVISSIE-----491
DB 428 ---DTSERSLNPSV-----LGLVFEKNGYKEGSPFTSPFITTSYCKESITITVL 475
QY 492 -----EOLKT-----KSPLDIKIIDNCGSGHGLFSLICLDYL 522
DB 476 DFNQKNVTECDKLEKKNYKNSYKEDKKRYQLQLLLTLRICDPVAGSGHFLVSLNEM 535
QY 523 TEKVWYELDKFEDVKKLEDEYRVIIIESE-----EYDVQDSISKE 563
DB 536 V-LIAVELGLIASLHR-----YSLTLENDEIITHYPTGTBIFNYKKPLRENDPHHIIQKE 589
QY 564 LV-LKRMLLKXCIYGVNDINPISVEITWLSWI-----NTPIFGTPLSFIEH 608
DB 590 LPELKKSIIENCFLGVNDINPNSCEITKLRLWIELLYKSYVIFBEGKNTNALET-LPNIDI 648
QY 609 HIKAGNALLGYTKDEFFDIVKKFESGFSLFKKRIKEITIL-----EDLY 654
DB 649 NIKCANSLI---SRFNLDLKK-----IPNKKIQEYKDLVAQYKDPNPLYNKADLI 701
QY 655 QKIKGINDT-----TKEDIEKS-KKIYKEYEBSKOIDNLRIFSLIKLYSLSFDKS 704

DB 702 NKIQDKNTFFSLTKDPKTKAELEKTIKHIKYN-----FFAL-----DDKS 744
QY 705 LNMEFSDIASVISLIENILGN-KTSSSEDK-----KIEKIRKL-----SSYKFFH 749
DB 745 L-----LDGLNYFIPSLFGLTKLSPKBEAEAFASYGRIRALRKLKDDALSQGEYQNAFE 798
QY 750 YGIEFPDIQE-----GFDIVIGNPWEKTKENETFEFFSKHPIPNYRKLGIKQNIIOEI 803
DB 799 WRFEFPEVLDDEGDFLGFDCIIGNPPY-----IRQEQ 830
QY 804 LSKDNHP-LSIEYNEEKNSIIAINNIIYKPDFKCTSGGDPNLFYFVTNFKLIIKEKGNL 862
DB 831 I-KDLKPLLEKQYQDFYNS-----TADIYTFYFALSFHLLREKGS 870
QY 863 TYLVPSAIWNSSRILRKHIFARYKLANIYQFENKGRFQVHSSGFKFAIFOLSNIKEST 922
DB 871 AFITSNKYARAKYGAKLREWLKTKTLVSYMELNALK-----VPESAAVDTSI 918
QY 923 SSFKAKFMIOSSDNILKEITRLDKS-KDAYKGIHNLINQIKLSPTQESIEFKDNEE 981
DB 919 IHFIKQPLKESDFKYBYPTPDODKDLKSTPYLSMKQNALSTESFIFANATLLDLRD--- 975
QY 982 FTLINKMFSKPSALGEGYIDFKKGLDPSIK-----NRKSLKKEC-----NN 1022
DB 976 -----KIEVSGTPLKMDIQINYGIKTCANEAFFIPTKREDEILNACKTOBERET 1026
QY 1023 KNL1-----FLYSGANIHOFSRFFEDKDAKESKLLWIDKE-----D 1060
DB 1027 ERLIKPILRGDKIKRYSYEWAGEWLINTHNGYTSNLK-SKIPPIDIEKYPATKAHLDAH 1084
QY 1061 LEKVUTK-----DNQOTERVFRATASNTNERTMTISTLSPGNCYCVNSIYINDEKTPISL 1116
DB 1085 YDTIATRCQDQPTPHLRNCAY--LEDFEKEKIVVPETSGQAYFIYENGSGIFLEKTVFMI 1142
QY 1117 -----YKCLFIISFNSFVDFLLRFRVDSNVL-----KSLYOCMPQPEEKEILSN 1164
DB 1143 VSDANLKLTLTALLSKLITFYFKNFCGCGCIIGKSGYQYNKHALEKIPQITERNOELA 1202
QY 1165 PLYLMLAKNTSLIARNDPENPKYLLYLEYFPRFDEKVKNLKLDKEDFFKEKENENN 1224
DB 1203 DKITCAEQILAKADPKANTQKL-----EKEIDA 1233
QY 1225 IIASLYSLAKEDFITL 1240
DB 1234 LVYQLYHLTDEBIKTI 1249

RESULT 4
US-09-738-626-4321
; Sequence 4321, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03

```

; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4321
; LENGTH: 1532
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4321

Query Match          6.7%; Score 435; DB 3; Length 1532;
Best Local Similarity 18.2%; Pred. No. 6.3e-16;
Matches 270; Conservative 255; Mismatches 507; Indels 452; Gaps 58;

QY 3 TNDIVTNNPNISLYKQSKDFTKKENINKLDFLILKIKLPSIDNTEAN-IESLLK 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 43 TEDATQSGP-----LTRSSNELQ-----LQHALSELDDATTAASLVASLG 85

QY 62 YIF-----EELNYSVEQKAGIEGVESRVDILLFENDKDKASFNKLKEAKGN 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 YGVPSARHAQRGSDTTISYSSWGNAGSVFLAATPAESFEENFR-----SLPLEPVAVD 140

QY 112 EPIPIEDILLIAEVKPTFSFDAKDKLSESDQLYRYLQYQKHGYLSNGKVRRLYDKS 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 KPQDITAAKLVGQI-----FLSDTPPAFVVITAGK-W----- 171

QY 172 KVLVGEK-----RYIEFNFKIEEKEEYKEQWFLVIYLIRKERYLKTSN-----V 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 -VVLARETWPLGRHLAIDISLVVERNDTKAQGEMQOTVVVALARENTERAADGTTWEET 230

QY 219 IEVEKQIAKEKEIIOKTKELIYERPDSDIVFKIARIYDKSFKLSGKEITOHILASIL 278
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 IEQREHAVKVSGLRSVRESIEILGNVLTTRYAEKELSTAE--IDGGELA----- 280

QY 279 EESIIFILRFFIAYIE---DNIDFKKLOENKLYRSSISFRYPFYDE----- 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 281 KQSLRYLYRILFLFAEASPELSILPTGTP-----YDEGYGLARLRELIL 326

QY 324 -----NTKKKLEYKIITIFNLJDK-----GSDAIFPVFNGGLFS 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 NPPVQHARTGTHL-YDSLGQLFTLVDKGHPHTDAKAVADAEAGEGLHFRNLSDLFL 385

QY 360 EDVKYLNNEGLLSISBIEBILVKMLPFEEKNKDEKFKVYRLDPSFGELYETLLEYD 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 386 PAATELIDRVG-LSNEALNKVLENLLSRVQSGKDRGFISYATLVGTQLGVYEGLSMT 444

QY 420 LR1A-----DTVHRIIEDGVY-----IRTEBELENKKYKNI-ATYLVKGN 459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 GF1AQEDLFVAPHGKADKGSWMLPVSKADEVPADSFIEVDQAPGGGVIVKRRHPRGS 504

QY 460 IYLTSLRDKKSGAVYTPDDLTFMVISSIEQLTK-----SPLDIKIIDNSCGSGH 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 505 FVFRQSRDRERASFTTPQVLTSTVTOAIELOAKRITTANDVLSLTICEPAMGSGA 564

QY 514 FLISCLDYLTQVMYELDKFEDVKKELDEBYRVIIESEEYVDQDSISKEL-VLKRMLLK 572
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 565 FAVEAVRQLAE-LYLEL-RQEELEQQIPAEGR-----AKELQVKAHIAL 607

QY 573 XCIYGVDPINISVEITMLSMINTPIFGTPLSFIEHHIKACNALLGYTKDEPDIIVKKKF 632
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 HQVYGVDLNSTAVELAEISLWDLTMNAEMDAPWYGLHRNGSLVGATSLYAP----- 661

QY 633 ESGFSLPKKRIKEIT-----ILEDIYOKI-----KG 659
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 -----SLNKKAWLTATPTRRLDDIAQAIDENKAPLNFNHHFLPSTGWGATADAKD 717

QY 660 INDTTKEDIEKSKYIKEYEES-----KDIDLN-----RIIFSLIKYLSLFDKSL 705
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 718 LKDLMAEIKELKSWRTSIRASLSKTKQIKQLNNLALRVETLWRFLVWRIRIAESQISRST 777

QY 706 NM-----EFSDSIASVLSLNIENLGN----- 725
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 778 TLMGQEPAEVSEVVREIQIDQLFGNIDGAYNRLRLVMDAWCALWFPLDVAETAEPHER 837

QY 726 -----KTSSEDK-----EK 734

```

```

Db 838 PALPDLDEWLATLTILGIDLPLKSKNENQIVLGPDTNWLAINDAEATDLGFSGALSFER 897
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 735 -----TEKIRKLSYYKFFHYGIEFPDI--QEGFDIVIGNPPWEKTFKNETEPFSKHI 785
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 898 VSANHPWINVARQAKQOQSFHWDLDFAHVFAKGFQDLQVGNPPWVRPDVNFEDLLAEHD 957
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 786 PNYRKLGIKEQNIIEKQILSKDNHPLSIEY-----NEEKSIINAINNIYKFPDFKCFPTS 838
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 958 FFWAVMSKPTQASKERQKFNHNPKSLSHVVSAGETVATSAILGSLVLYP-----HLK 1012
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 839 GGDPNLFRYVFTFNKLKEKGNLTYLVPSAIWNSSSSILRKHIIPARYKLANIYYOFENK 898
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1013 DQRPDLRYGFMETKWSNASPAGASILIHDESHFTEKKAAPLRGAYERLRRH--WQFINE 1070
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 899 KXPKDVHSPKPAIFQLSNIKESTSPKAKFMIQSSDNILKEITRD---LKDSKDD--- 951
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1071 LILFDVHDLVKGVHVGHYGAQESI-NFLJSAASLYHPQTVLDSLDDHGGSNLPGLKDDNGN 1129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 952 ----AVKG-IEL-NINOIKKLSPIOESIIEFK-----DNEEFTLINKMFS-- 990
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 WDRPHKORIQLVNAIDL---TWKLSILEDEQTPVLDTRMVTYVNTTEAAVLEKLASAP 1185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 991 KFSALG-----EGYIDFKKG-----LDPSIKNRKSLKE 1019
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1186 RIKELGLQFSSGWNETTDKKGFDVWGVPASWSDAILOQPHLGVAATPMIKQPNPTMK- 1244
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1020 CNKKNLILYSCANTHQFNSRFF-----EDKAKESSKLLWI--DKEDLEKVLTKD 1068
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1245 -HNQD-----WSEIDFEAIPANFIPATAYQPDRTQTPYDADYGTWTFGDKQ-----VPVAD 1295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1069 NOYQTERVYRAIASNTNERTMI-STLSPGNCVCNSIYINDEKTPISLVKKLFIISFN 1127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1296 ----TFRIAREMAATGRTVYPSVPEG-----ANHVTVAASRSNLKILVGAQLG 1347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1128 SFVDFLLRRFRYDVSNLKSLYQCPMPQ--PEEKILSNPLYLN 1169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1348 AILSDYFARSSGSSHIFNDIVRKPLPNFTSLEKQFARTYLRLN 1391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
US-09-882-227-388
; Sequence 388, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 388
; LENGTH: 1279
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-388

Query Match          6.4%; Score 419.5; DB 3; Length 1279;
Best Local Similarity 21.1%; Pred. No. 3.8e-15;
Matches 308; Conservative 209; Mismatches 454; Indels 487; Gaps 70;

QY 68 NYSVEQKQAGQIEGVESRVDILLFENDKD-----KASFNKLKEAKK----- 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 NYPNOE-----QLKAFETAFDAFLETTNQENENHQNDAFNDDLKGVFKYKPKTKIDSTI 67

```



```
Db 193 FLQOI-ENWREKLAITAIONTE-----LGEEDVNFIVQRLNLIIFLVCEDRTIEKY- 245
Qy 304 LQENKLYRSSISPRFFYDENTKKLEYKKIITENLNDKGSDAIKFPVFGSLFSEDKV 363
Db 246 -----ETIKSIKNYBELKDLFOKSDR-----KFNGLFD----- 274
Qy 364 KYLNNEGLASISIEBILVKM--LFFEKNIKDERFVKYSRLDPKPSFGELYETLLEYDL 420
Db 275 -FIDDTLLLEVEDSNVLEIFSDLYFPQSP-----YDFSVDPTILSQIYERFLGQRI 327
Qy 421 RIADTTVHRIEDGVVYLIRTEBELENKVNKIATYLGNIYITSRSLDRKSGAYYTPDD 480
Db 328 -----IIESS-----GTFHITESPEVAASNGVPTPKI 355
Qy 481 LTDPMV---ISSTEEQKTKSPDIKIIDNSCGSHFLJSLCDLYLTKYVWYLDKPEDVK 537
Db 356 IVEQIVKDTLTPTEGKKNELCNLKIADICGSGTFLISSYDFLVKRV-----ME 406
Qy 538 KELDEBYRVIIESE-EYDVQDSISKELVLRKMLKKCIYGVNDINPISVEITMLSWINT 596
Db 407 KIIIEEN---IDSDLVYETEGLILTLAKRNILENNLFGVDVNVYAVEAEFSL----- 458
Qy 597 FIFGTPLSFIEHHIKAGNALLGYTDEFFDIVKKFESGFSLPKKRIKIIITILEDIYOK 656
Db 459 -----LLKLE----- 464
Qy 657 IKGINDTKEDIEKSKIYKEYESKDIDNLRIFIISLIKLYSPDKSLNMFSDIASVI 716
Db 465 --GENEASVNN-----FIHEHE-----DKIL-----PDUTSI 490
Qy 717 SLJENTILGNK-----TSSDKEKIEKIRKLSSYYKFFHYGIBFPDIOE--GFDIVIGNP 768
Db 491 KCGNSLVNDKFFEFMESLEDEILKAPN-----FEWEEFPDIMANGGFDALIGNP 543
Qy 769 PWBKTFFNETEPFSKHIPNRYKLGIXEQNIIOEILSKONHPUSIEYNEEKSSIINNI 828
Db 544 PY-----VRIONM-----KKYSPEIBIYQSKDE----- 568
Qy 829 YKDFDKFTSGGDPNLPFRYVTFN--LKLIKEGNLTLYLPSAIWNNESSRILRKHIPAR 886
Db 569 -----YTVAKETVDKYFLPTERAILLNPTGLLGYIIPHFFITKGGKELURKPIAEK 621
Qy 887 YKLNLYIQPENKRFKDVHSSFKFALPQLSNIKESTSSPKAKFWIOSSDNILKEITRDLK 946
Db 622 HQISKIINFQVTPFGGAYTALLIIOANKM-----AQPKYKVNISAEYL----- 669
Qy 947 DSKDD--AYKGIELNINQIKGLSPIQESIIEPKDNBEFTLINRMFKSPSALGEGYIDFKK 1004
Db 670 DSEBNTCVYSSEKYNSDPMWIFLSPTEAV-----FTKfte--AQFEKLGb-ITDISV 718
Qy 1005 GL-----DPSIKRKSLLKE-----C-----NNKQLI 1026
Db 719 GLQTSADKIYIFIPENSTDTYIFNYKGRYIEKSIICCPAIVDLSPGSPESIQGNAQMI 778
Qy 1027 FLYSGANIHFNSRFFDKAKESSKLLWIDKEDLEKVLTKDQYOTERVFYRAIASNTN 1086
Db 779 FPY---EIRDEAXLLEETLENDYPLAWNYLNEFKALEKRSLOGRNPKWYQGRSQSL 835
Qy 1087 ER-----TWISTLSPGNCVNSIYINDEKTPISLYKKLPIISFNSFVDFLLRR 1137
Db 836 SKPHDKELIWTVLATKPPYVLDNRNLLFTGGNGP---YYGLINQSIYSLHYF----- 886
Qy 1138 FVDSNVLSKCLYQCPMPQPEKEITLSPNLYLNLAQNTSLIAKNDPENFKYLLYLEYFKF 1197
Db 887 -----LGLSHPVIESMYKARA-----SEFRG-----SYISH 913
Qy 1198 DKEKVNKI---LKLKDEBFFK-----EK-----ENENN----- 1223
Db 914 GKQFIEKIPKIDFDQDEVDKNTVTVTVEKLIITTDRIKSESNGPRRMLRRLLDAL 973
Qy 1224 -----FIASLYSLAKEDFTLLGD 1243
Db 974 SNOLIOVINELYNISDBEYTVLND 998
```

RESULT 7

```
US-10-732-923-3352
; Sequence 3352, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3352
; LENGTH: 1441
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
US-10-732-923-3352
```

```
Query Match          5.4%; Score 354; DB 5; Length 1441;
Best Local Similarity 20.2%; Pred. No. 2.3e-11; Indels 498; Gaps 76;
Matches 312; Conservative 255; Mismatches 479;
```

```
Qy 10 NNPNISLYKQLSK-----DFIKKE-----NIN-----KLDPFILIK-NK 43
Db 54 NELKLEKEQINKEYEKELNEKCEFIKREMEKELKEKELNINENKINNKEIITLKKEK 113
Qy 44 LPSIDONSTEANIESLLKYIFELNYSVE-----QQKAGQIEGVESRVDILL 90
Db 114 LNDIENEYIEKNKKEK-----EKLNYEITNIMQSLDKLSCEIQEKENLEKINKVY---- 163
Qy 91 FENDKDKASFNKKLEAKKQNNPEPIEDILIIAEVKRPTF-----SFDKDKLKESEDQL 145
Db 164 ----VEKENNRELKEFMEKKEKEI--IESLNTKIDDKKNAYEKLNTFEBKRMIMLDS- 217
Qy 146 YRLNQYQHYGILSNGKVRWLYDKSVLYGKRYIEFNFXKIEEKEEY---KEQSWFVL 202
Db 218 --KLIEKEKFFI-----KKEKLEKENEYIMEKLDKDIENKEEHFKNKEE--- 260
Qy 203 FYILRKERYLKYTSNVIEVEKQIAK---EKIIOQTKLKEI---LYERPDDSVIFKIAKN 256
Db 261 --FQKKEKFINLENELNKLKSLSKNAQOMIYKMEIKDLSQSLVEKERE--IFBI-KN 315
Qy 257 IYDKEP-----KLSG-----KEITQHILASILEESIIFILRIFFIAYIEDNDIFK 301
Db 316 EYDDKINNMMKNKLSNPNDDNDNTVKCSSEIINKKIEEAVNKLK--EKEMELNEIHK 372
Qy 302 KILQENKLYRSISFRFFYDENTTK-----KLEYKKIITI-----FNLLD 342
Db 373 KYNLEIEKIKNEINEKEKELEQNKKKHNTIENDLTKEIQIREKKIEDVKEEYKIELSKLD 432
Qy 343 KGSDAIKFPVFGGLPSE-----DKVYLNNEGLLSISEIEELVQMLFEEKNIKDEKF 397
Db 433 SEKQNIK--IENNELNNEVNSLNNE-----VNSLANNEINLNNDK- 477
Qy 398 VKYSRLDPKSPGELYETLLEYDLRIADTTVHRIEDGVVYLIRTEBELENK--KVNKIATY 455
Db 478 -QTLKKNKLLNDLIN--LKNEINNSDNKNQMKED---IIMLNSELEGKCVVIDIEKK 532
Qy 456 LKGNILYTSRSLDRKSGAYYTPDDLTDPMWISSIEBQKTKSPLDIKIIDNSCGSHFL 515
Db 533 YKNEIIFPEKKLEKEN-----YDDLNDENILNNDISVKEKEFIEMK----- 575
Qy 516 ISCLDYLTBKWYVELDKPFDVKBELDEEYRV-----IIESEEDVQDSISKELY 565
Db 576 ----EFYENKINLYKNFBEKTKYENELNILKLTNNQODLIKQIDELNIQKLKTEBEY 631
Qy 566 LKEMLLKXCICYGVNDINPISVEITMLSWINTFETGTP--SFIEHHIK-----AGNALL 617
Db 632 LQ-----LYDNNIKOMFKSICAKINTTYSENIKGSDLVNFVAYIKHRDESFPDANSES 684
```


Db 1730 LDTE---KXNNQMSIELDEKKOMHORRIELDEKKOMHMMNIELDEKKINKINILL 1786
 QY 819 ----KNSIIAINNIYKDFPKFTSGGDPNIFRYFVTNKLKIEKGNLTLYVPSAINNESS 875
 Db 1787 EENKNKIIQINS-----CLOKEKEIN-----ELLNIOLEDKKNNIQEL---NIKLEDE 1831
 QY 876 SRLRK-HIFARYKLANVIYO---FENKGRFKD---VHSFKAIFQLGNIK---ESTSS 924
 Db 1832 KCLNEQINTVLENNKINNIQELNKLDEKKLNEQINITLENKINNIOELNILEDEKKON 1891
 QY 925 FKAQMIQSSDNILKEITRDLKDSKODAYKGIELNINOIKKLSPIQESIEFKONBEFTL 984
 Db 1892 QQINIQLEBKNKIIQINSHLKKEKE---INEELNILENNKONNIQETIKLEBQKK--- 1945
 QY 985 INOMFKSALGEGYIDFKGLDPSIKNRKSLKCEKNKILFLYSGANIHOFSRFPED 1044
 Db 1946 INE-----BQYIQLEK--DKEIIN--SMVMEKEKII-----NN---EI 1978
 QY 1045 KDAKSSKLWIDKLEKVLTKDONOVOTERVYFRAIASN--TNERITMISTLSPGNCYCV 1102
 Db 1979 KQKLEKEK-----KONDQLVIHLENEKQANKKLNILLDQNKINIEL----- 2020
 QY 1103 NSIYINDEKTPISLYKKLFIISIFNSFVDFLLRRFVDSNLKSLCYQCPMPQPERKEIL 1162
 Db 2021 -NIQVEQEK-----LJNNEIIVQLAKENEENKINSLL-----EQON-- 2056
 QY 1163 SNPLYNLAKNTSLIAKNDPENFKYLLYLE-----YKFDKEK-----VNKILKLOK 1210
 Db 2057 -----GLNKKVTLQLEKEEENGKGLQLENEKQBNGLRPELENEKKDIANLILQOE 2110
 QY 1211 EDE-----FPKEKENENNFI 1226
 Db 2111 EKENTKNVMQMDKEKTKKNVMV 2134

RESULT 9

US-10-732-923-3340
 ; Sequence 3340, Application US/10732923
 ; Publication No. US20050108791A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52796)C
 ; CURRENT APPLICATION NUMBER: US/10/732,923
 ; CURRENT FILING DATE: 2003-12-10
 ; PRIOR APPLICATION NUMBER: 10/310,154
 ; PRIOR FILING DATE: 2002-12-04
 ; NUMBER OF SEQ ID NOS: 24149
 ; SEQ ID NO 3340
 ; LENGTH: 1939
 ; TYPE: PRF
 ; ORGANISM: Plasmodium chabaudi
 US-10-732-923-3340

Query Match 5.18; Score 335.5; DB 5; Length 1939;
 Best Local Similarity 20.18; Pred. No. 3.8e-10;
 Matches 304; Conservative 248; Mismatches 502; Indels 457; Gaps 75;
 QY 18 KQSKDFIKKENINKLADF----FILIKNKLFSIDNSTEANTIESLLKIYFEELNYSVEQ 73
 Db 115 KQINKEYEKE--LNEKSEFIKQWELKKEKELNINLENKINKEIITLKEE----- 165
 QY 74 QKAGQIEGVRVDILFENDKDKASPNKKKEAKQONEPIPIBIDILIIAEVVRPTSPFD 133
 Db 166 ----KLNDIESE----VIEKKEKEKLYEVNTIKMSLDKLTCB-----VQEKK-----D 207
 QY 134 AKDKLK-----ESEDOLRYLNOYOKHY-----GILSNKQVRLYDKSKVLYGKER-Y 180
 Db 208 NLEKINKKVIKKEKNNL-RELKBFKKEKNEIIESLDGTINDKK--NAYEKLEISFEERKM 264
 QY 181 IEFNFXKIEEKEEYKEQEWFLYTLIRKERYLTSN--VIEVEKEQIAKEKEIIOKTLK 238
 Db 265 IEMLDKSLIEKEE-----NFANKQAKLEKEKEIIEKLDIESREKDKPSKEEK 313

QY 239 EILYERPDSDIVFKIAKNIYDKE--FKLSGKEITQHILASILEESIIFILRIFF----- 290
 Db 314 FASMENELNTLSKDSKACQMEVYKLEIKDLSQSL---VEKEREIIFELKNEYDDKINNM 370
 QY 291 ----IAYIEDNDI-----FKILOENKLYRSSISFRYFFYDENTKKL----- 329
 Db 371 KEKLSINDKIGDINTVLHSEEBKINKLKEKETELNEIHKYNLEIETIKNELNEKEEEL 430
 QY 330 --BYKKIITI-----FNLLDKGSDAIFKFPVNGGLFSEDKVKVYNNNEGLSISIEBI 380
 Db 431 EKNKKAHTVEVNTLTKLEKKTEDAK---EG---HKNELNELNQ---LSKLNK- 478
 QY 381 LVKMLFPEEKNIKDKSEF-----VKYSRLDPKFSFGELYETL-----LEYDLRI 422
 Db 479 -----EKONIKNTELNELDKISINSEVNILNKDKQTLGNDIKTLNDLNNLNKNEINT 531
 QY 423 ADTTHRIIEDGVILIRTEEELENK--KVNKIATYLGNIYLTSLRDKSKSGAYTPPD 480
 Db 532 SDNKNKMKED--LAMLNEEMEGCVVIDEIEKKYKNEIFMLEEKLKEKENYA-----D 583
 QY 481 LTFDFWISSIEBQLTKSPLDIKIIDNSCGSGHFLISCLDYITEKVKWYELDKFEDVK--- 537
 Db 584 LND-----EISILRNSI-----YVKEKEFIEKMEFYENKINL 615
 QY 538 --KELDEEYRV-----IIESESEYVDQDSISKELVLKRMLLKXCICYGV 578
 Db 616 FKNFEEKKNYENELNSURLKYDNEQGLIKQIDELNIOKLKTEEKYLO-----LYND 668
 QY 579 DINPISVEITMLSLMINTPIFGTPL--SFIEHHIK-----AGNALLGYTKDEFPDVIKKKF 632
 Db 669 NMHMFRICTIDMPYSENIKGSDLVDVFTAVIKRRDESSDANPDPTTHKEMVAELEKRH 728
 QY 633 ESGFSLFKRIKEITIL-----EDIYQIKGINDTTKEDIEKSKKIY 675
 Db 729 AAIVAELEEKHEEIAKLGEHKEVVLRLGEQHEKEETIILEBKHKDVVTKLGEQHKENI 788
 QY 676 KEYESKDIENLRIIFSLIKLYSLSPDKSLNMFSDIASVISIENILNKTSSDEKEKI 735
 Db 789 KLEEEHKOV-----VTKLDQYKEEIAKUEEHAVVVAELEEK---HKLGEHKEMV 837
 QY 736 EKIRK-----LSSYKFFHYGIEFPDIOEGFDIVIGNPWEKTKFNETE----- 779
 Db 838 DSLEKHEHADFVGLKEKHK-----AETAKLEEG-----HKSEMNEVEKRAHDFVE 882
 QY 780 --FSPKHIPNRYKLGIKEQNII-----KQEI--LSKNHPUSIEYNEE 818
 Db 883 GLEEKHKAETAKLGEHREVAVAGLEEKHEVVAELEEKHEEIAKUEEHKEVMAELGEK 942
 QY 819 KNSIIAINNIYKDFPKFTSGGDPNIFRYFVTNKLKIEKGNLTLYVPSAINNESSRI 878
 Db 943 HKEWVA-----GLEAKHNLBEGHEMV-----AELEKRAHDLV-----AVLEEQHKA 986
 QY 879 LR-----KHIFA-----RYKLNVIYOPENKK-----RPFKVH----- 905
 Db 987 IKLGEHEKEVAGIEBEKYVEALKAEHKKDVVTKLGEQHEKIEIAKLDGHEKVVNEVEK 1046
 QY 906 -----SSPKFAIFOL--SNIKESTSFKAKFM-----IOSSNLIKIEITRDLKDSK 949
 Db 1047 KNASLLNMLEENHNKEMIKLKEHKEBSASDLVEKLYQKOBENVKNSNNKIBELTNVINDL 1106
 QY 950 D-----DAYKGIENINOIKKLSPIQESIEFKONBEFTLNNKMFSPSALGEG 998
 Db 1107 DSMCYKQOILEVEVEKREYN--EEINKLKIYQN---EMKMDNKKILEK-----EN 1153
 QY 999 YIDFKGLDPSIKNRKSL-LKECENNKLIFLYSGANIHOFSRFPEDKDAKESKLWID 1057
 Db 1154 EI---KLNKCLSNYKVFETKENTYKN-----SEMVVNENKERIIVDSVCKEN-----IS 1200
 QY 1058 KEDLE-----KVLTKONQYOTERVYFRAIASNTNERT--MISTLSPGNCYCVNIIYINDEK 1111
 Db 1201 ESDVEGKGGLNKWTLSLKKERNIF--SINDNKNESELVDI-----IKSAYIN--- 1247

QY 1112 TPISLYKLP-----IISFNSFV-FDELLRRFVDSNVLSKSLYQCPMPQPE 1157
DB 1248 -KIEMTKEDNGKNIEDKNKILDSNELINLENKNVLTDDNNLKEBIKDNKN 1306
QY 1158 EKEILSNPLYNLAKNTSLIAKND-----PENFKYLLVLEVPKPKKEKVKNI 1205
DB 1307 EKEKNTEILNL--NDDIIFLKEISEWKDEBEKLTENIK-----LKNDEIQINKE 1357
QY 1206 LKLDKDEFPKKEKENNNFIASL-----YSLAKEDFTLLGDFAKLNKKKGEDYISLSIK 1262
DB 1358 YKIKEENMLKPFENINE--VTSLNQIETKMKBELKNVYLLAEKR-ETNMS--IS 1412
QY 1263 GYDNYLLNKKI 1273
DB 1413 NDNKLVENNI 1423
RESULT 10
US-10-732-923-3354
; Sequence 3354, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3354
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; NAME/KEY: unsure
; FEATURE:
; LOCATION: (1)..(1527)
; OTHER INFORMATION: unsure at all xaa locations
US-10-732-923-3354
Query Match 5.0%; Score 328.5; DB 5; Length 1527;
Best Local Similarity 21.3%; Pred. No. 7.1e-10;
Matches 247; Conservative 210; Mismatches 386; Indels 319; Gaps 61;
QY 1 MKTNDIVKTNPNISLYKQSKDFIKKENINKLKDFLILKNKLFSDNSTANIESLL 60
DB 544 MDENPIILDSNTIN-----KNNIIDAESDKLVQNLDMIKIL--QDDKNK--NIDTII 594
QY 61 KYIFBELNVSEQKAGQI-----EGVESRVDIILFEN-----DKOKASFN 101
DB 595 KNESEKIT-SLLEKTSMBILCNKICQLKKNEDVEDKNDTIREETHQINDKKEBI 653
QY 102 NKLKAKKNEPIPIEDILLIAEVKPTFSFDAKDKLKESEDQLYLYNQYKHGILSN 161
DB 654 NNLLYV-----INIKNKIKKELEAFTNSFNLKLMKNDBI--INIKRNVNFLT 704
QY 162 GKWRLYDKSKVLYGKRYEFNFXXKIEEKEEYKEQWFLVLYLIRKERYLXTNV-- 218
DB 705 -TINKLEEKRL-----KYAEFK-----EKEKYKE-----LIMKKEELLIXQINK 746
QY 219 --TEVEKQIAKGBIQQ--TLKELYERPDSIVFKIAKNYDKFKLSKKEITQHI 273
DB 747 KNITFPNQBMQLKDEIQIELTKETKIEBLKQD-----FEKTQINMENLEME- 795
QY 274 LASILEESIIFILRIPIAVIENDI--FKKILQ--ENKLYRSSISFRYFFVD--ENTK 326
DB 796 -----RES-----FINNNLEINMKKELEEKYETQIKETMKYQIKKEIEKTK 842
QY 327 KLEYKXIIITIFNLLDKGSDAIFPVPNGLFSFEDVKVYLNNGLLSISIEBI----- 380
DB 843 QNAEQN-----FN-----SKPEKYKEML-EKNKNDFINNLIIEKNNEIESFKNDIEQ 888
QY 381 --LVKMLFPBEKNIK-----DEKFVKYSLRDPKSGELYETL-----LEY----- 418

DB 889 KKFEMEKKEBNEKLLQNNFENMKNHFIIEQKNKHIEIKKYEELIKNEIEYLYKEEMKK 948
QY 419 -----DLRIADTTVHRIIEDGVYLIRTEBEEL--NKKVKNKIATVLKNGIYILTSRS 466
DB 949 NKIQEITENVELKAD-----EKNKHIDDMKKELEINYNVEINNLKEKINAEQTQAEM 1001
QY 467 LBRKSGAYTDPDLTDFWVI--SSIEEQK-LTKSPLDIKIID--NSCSGHFLISCLD 520
DB 1002 LETQ-----NEVEQKLEIQORNEYEQOLEIRRNEMENKLEIQRKESYDLGSKKSEIQ 1054
QY 521 YLTKVWYELDFEDVKKELDEEYRVIIESEEDVDODSIS-----KELVLK----- 568
DB 1055 ILT-----DKIEKLKMEIE-----IINKKIDLEHNSILNDRERILNQNKULEDN 1101
QY 569 MLLKXCIYGVINPISVEITMLSLMINTFIFGTPLSFIEHHIKAGNALLGYTKDEFFDIV 628
DB 1102 LKNNEIYNKINILNDEKLE-----KEIEHIKENGAKESQDIREQFADLL 1149
QY 629 KKKFSGFLFKRIKEIITIL-----EDIYQIKIGINDTK-----EDIEK 670
DB 1150 QBEIDR-----IKESKEKVNIIQOYNEINEEYEEKKEKYNOLLEKAVSNKQTEKCEB 1205
QY 671 SKKIYKEYES-KDIDNLRILFSLIKLYSL-----SPDKSLANWESDIASVISLIE 720
DB 1206 NIQKINEYEDMIKMLENQTENVLTKIQBELNEDFLKKEAFNEKNDLLKNYEHAIENK 777
QY 721 NI-----LGKNTSSDEKIEKIRKLSYKFFHYGIEFFDIOQEGDIVIGNPPWEKTKFNE 777
DB 1266 HIKEQLENFTNS-NEEKISQIKN-----QYETQIKEMQIHFDL-----NETKNKN 1311
QY 778 TEFFSKHI--PNYRKIGIKENIIOBILSKONHPLSIBYNEEKNSIIANNIYKDFPKC 835
DB 1312 NEMLEDIVKEQNDKEQSIQHNSVNTLVTK-NEQLNLNITETIKNEL--SNFOEKYD--- 1365
QY 836 FTSGGDPNLPFYFVFNKLKIKEKNLTYLVPSAINWESSRLRKHIFARYKLYIYQF 895
DB 1366 -----KLVKENNHLK-----NEBQTFPLDKILDLENVKNHLMEV 1399
QY 896 ENKCR---FKDVHSGFKPAIFQLSN-IKESTSFKAKFMIOSSDNLILKEITRLDKSKD- 950
DB 1400 TEKEREINKNV-----IKSLQIQINEQTKLYK-EYTDLKDDEL-QLKKQKNSIDS 1452
QY 951 ----DAYGIELNINQIKLSPIQESIIEFKONEEPTLINKMFKSFGSALGEGVIDPKGL 1006
DB 1453 NVPIDANSSINNALLEENKLLQNEVISTKNDTMTSLDNLTKRLS-----PIESK--- 1504
QY 1007 DSIKNRK---SLLKECNKNL 1025
DB 1505 ---IKEEYGNDIIOKANDLQ 1523
RESULT 11
US-10-732-923-3351
; Sequence 3351, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 3351
; LENGTH: 1985
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
US-10-732-923-3351
Query Match 5.0%; Score 324; DB 5; Length 1985;
Best Local Similarity 19.7%; Pred. No. 1.8e-09;


```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...740
; SEQUENCE DESCRIPTION: SEQ ID NO: 7398:
US-10-335-977-7398

Query Match          4.9%; Score 318; DB 4; Length 740;
Best Local Similarity 24.1%; Pred. No. 1.1e-09;
Matches 176; Conservative 99; Mismatches 240; Indels 214; Gaps 35;

QY 22 KDFIKKEN-----INKLKDFILIRKSLFSDNNSTEANIESLLK--YIFELNYS 70
DB 54 KVFIKYQEPKPKETIENFEKEINSLENVPRQDDDEEQNEINFLKNAVGYDQNTY- 112
QY 71 VEQQKAGQIEGVSRVDILLFENDKDKASFNKKLKAQKNNKPEIP-----IEDI- 119
DB 113 ---EKVDSVIYVDKEVRVLI-----EVKALNNKTEFPKKNRENPLSKAFQCMVLYFLEEIS 164
QY 120 -----LIIAEVVRPFSPPDAKDKLKESED-QLYRYLNYQYKHVGHLSNGKVRWLYD 169
DB 165 KKNNSLKHITIICNA-HEFFLPCKDOLLSLKEDKRIKKFYKNYAKKEGT-----D 213
QY 170 KSKVLY-----GKRYIEFNFEXKIEBEKEEYKEQEWFLVFIYLIRKERYLKTS 216
DB 214 SSKPKFYKLEQPLQEDFQGLRYTHFNL-----SDFKX--LPLIYQVLSQVLLKQK 265
QY 217 NVIEVEQOIAKEKEIIQTKLEILYERDD--SIVFKJA--KNYIDKFKSGKEITH 272
DB 266 RTLDAN-----TLNKDYEBELLYILGLEQNDKILIKPSRTKNSLSGALKKERYKNLDB 321
QY 273 ILASILEESIIFILRIFFATYIEDNDIFPKKILQENKLYRSSISFRYF-----FYDENYK 326
DB 322 EYVALL---IANNRILFLRLLESLLISFEHFEKERSFLTGNFKDFNALNTLFFEVLA 378
QY 327 KKELYKKIITIFNLDKGSDAIFPVFNGGLFSE-----DKVKYLNNEGLLSISEIEE 379
DB 379 KNSERLPEIKENKILG-----KIPYNSLFPDKTPLELKGHEIKLLDNKKL-----E 425
QY 380 ILVKMLPFEEKNIKDKFKVYKSLRDPKSPGELYETLLEY---DLRIAD-TTVHRIEDGV 435
DB 426 IYKNSVLKXHENYQDK-----ASLPLLEYFFKFLRVYDFTTTPKIDKWT 471
QY 436 YLIRTEELEENKWKNIATYLGKNIYLTSLRDKKSGAYTTPDDLTDFWVSIIE- 491
DB 472 --DTSERLNPVS-----LGLVPEKNGYKEGSFYTPSPITSYNCKESITTVIL 519
QY 492 -----EOLKT-----KSPLDIKIDNSCGSGHFLISCLDY 522
DB 520 DFNQKNVTECDKLEKKNYKNSYKEDKRYQLQLLLTLRICDPVAGSGHFLVSLN- 577
QY 523 TEKWV--YELDKPE-----DVKKELDEYRVIIIEES-----BEYDQVDSISKE 563
DB 578 -EMVWVAYKGLTASLYRYDYLKLEND--ILIHPTGTGEIFNYTIPHSENDPHHIOKE 633
QY 564 LV-LKRMMLKCYGYVDINPISVEITMLSWI-----NTFIFGTPLSFIEH 608
DB 634 LPELKKSIIENCLFGVDINPNSCEITKLRLWIELLYKSYVYIFBEGKNTNALET-LPNIDI 692
QY 609 HIKAGNALL 617
DB 693 NIKANSLSI 701
```

RESULT 13

US-10-954-924-7

; Sequence 7, Application US/10954924

; Publication No. US20050095256A1

; GENERAL INFORMATION:

```
; APPLICANT: Bujard, Hermann
; APPLICANT: TOLLE, RALF
; APPLICANT: PAN, WEIQING
; TITLE OF INVENTION: Recombinant Process for Preparing a Complete Malaria Antigen GPI
; FILE REFERENCE: GRUB-003DIV
; CURRENT APPLICATION NUMBER: US/10/954,924
; CURRENT FILING DATE: 2004-09-29
; PRIOR APPLICATION NUMBER: 09/269,874
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1602
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-954-924-7
```

```
Query Match          4.8%; Score 316; DB 5; Length 1602;
Best Local Similarity 19.7%; Pred. No. 3.9e-09;
Matches 295; Conservative 229; Mismatches 481; Indels 490; Gaps 73;

QY 2 KTDI-----VTNNPNISLYKQSKDFIK-----KENINKLKDFILIRKQLF 45
DB 172 KLVDCANDYCOIPFNLKIRAMELDVCLKLVGYRKPDLNIIKDNVGMEDY--IKNK- 227
QY 46 SIDDNSTEANIESLLKIYFELNYSVEQQAQIEGVSRVDILLFENDKDKASFNKKL 105
DB 228 -----KTIENNELI-----EESKTTIDKNKNTATKEEKKK---LYQAQYDLSIYNKQLE 274
QY 106 EA-----KKNNEPIEDIL-IIAEVKK-----PTSFDAKDKLKS 141
DB 275 EAHNLISVLEKIDITLKKNE--IKELDKINEIKNPPANSNGTPTNTLLDKNKIEEH 331
QY 142 E-----DQLYR-----YLNQYQKHVGHLS-----NKKVVR 166
DB 332 EKEIKEIAKTIKFNIDSLFTDPLELEYILREKNKNIIDISAKVETKESTEPNPGVTYP 391
QY 167 L-YDKSKVLYGKRYI-----EFNFKIEEKEEYKEQEWFLVFIYLIRKERYLKTNSVIE 220
DB 392 LSYNDINNALNELNSFGDILNPFDTKBPKNIYTDNE-----RKKFINEIKETIK 442
QY 221 VEKBOIAKEKEIIQ---KTLKEILYERDDSVIFKIAKNIYDKEF-----KLSGK 267
DB 443 IEKKKIESDKSYEDRSKSLNDITKEYE-----KLNIEIYDSKFNNNIDLTNFEKMMGK 496
QY 268 -----EITQHILASILEESIIFILRIPIFIAYIED---NDIFPKILQENKLYRSSIS 315
DB 497 RYSYKVEKLTHTNTPASVENSXHNLEKLTALKWMEDYSLRNIVBEKELKYKQLISKIE 556
QY 316 FRYFPYDENTKK---KLEYKKIITIFNLDDKG-----SDAIKPPVFNGLFSEDKVKYVLNN 368
DB 557 NEBITLVENIKKDEEQLEPKKITDKNKPEDEKILEVSDIVKVVQV----- 600
QY 369 EGLLSISETIEELVKMLPFEERNIKDEKVPKYSRLDPKSPGELFYETLLEYD---LRIADT 425
DB 601 QKVLLMKNKIDELKTKQILIKONVELKHNHVPNSYKQENKQEPYLIIVLKKKIDKLVFMP 660
QY 426 TVHRIEDGVYLIRTEELEENKVNKIAIYLGKNIYLTSLRDKKSGAYTTPDDLTDFM 485
DB 661 KVESLINEBKNIKTEGQSDNSEPS-----TEGEITGQATTKPGQAGSALEGDS----- 710
QY 486 VISSIEEQLKTSPLDIKI-----IDNSCGSHFLISCLDYLTEKW----- 527
DB 711 VOQAQEQKQAQPPVPVPVPEAKAQQVPTPPAPVNNKTEN-----VSKLDYL-EKLYEFLNT 765
QY 528 -----YELDKFEDVKKELDEYRVIIIESEEDVDQDSISKELVLKMLKXKCIYGVNIN 591
DB 766 SYICHKIYLVSHSTWNEKILKQYKITKEESKLSCDPLD-----LLFNQI-N 812
```


GenCore version 5.1.6
Copyright (C) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 20.4381 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-16
Perfect score: 6538
Sequence: 1 MKTNDIVKTNPNISLYKQL.....SSLIQYDNYLLNNKIFTHK 1277

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	284.5	4.4	2665	US-11-124-368A-214	Sequence 214, App
2	269	4.1	2710	US-11-051-453-41	Sequence 41, Appl
3	266	4.1	2668	US-11-124-368A-215	Sequence 215, App
4	261.5	4.0	1786	US-11-196-400-3	Sequence 3, Appli
5	224	3.4	1652	US-10-995-561-663	Sequence 663, App
6	224	3.4	1938	US-10-995-561-662	Sequence 661, App
7	224	3.4	1938	US-10-995-561-662	Sequence 662, App
8	222	3.4	1169	US-11-077-550-20	Sequence 20, Appl
9	219.5	3.4	1954	US-10-995-561-660	Sequence 660, App
10	219.5	3.4	1972	US-10-995-561-664	Sequence 664, App
11	219.5	3.4	1972	US-10-995-561-666	Sequence 666, App
12	219.5	3.4	5024	US-10-793-626-2594	Sequence 2594, App
13	211.5	3.2	1404	US-10-878-556A-169	Sequence 169, App
14	206.5	3.2	885	US-10-793-626-1660	Sequence 1660, App
15	206.5	3.2	1070	US-11/062	Sequence 4, Appli
16	206.5	3.2	1095	US-11/062	Sequence 7, Appli
17	205.5	3.1	2367	US-11-051-453-42	Sequence 42, Appl
18	198.5	3.0	1614	US-10-821-234-903	Sequence 903, App
19	192	2.9	1147	US-10-615-668-5	Sequence 5, Appli
20	190.5	2.9	976	US-11-155-288-20	Sequence 20, Appl
21	188	2.9	1960	US-11-069-834-48	Sequence 48, Appl
22	187.5	2.9	1216	US-10-873-528-12	Sequence 12, Appl
23	182.5	2.8	1976	US-11-069-834-54	Sequence 54, Appl
24	180	2.8	1315	US-11-077-550-141	Sequence 141, App
25	178.5	2.7	708	US-11-196-475-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1

US-11-124-368A-214
; Sequence 214, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 214
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-368A-214

Query Match 4.4%; Score 284.5; DB 7; Length 2665;
Best Local Similarity 20.0%; Pred. No. 2.6e-07;
Matches 291; Conservative 224; Mismatches 454; Indels 489; Gaps 73;
QY 1 MKTNDIVKTNPN-----PNISLYKSLKDFIKENINKL-----KD 35
1266 INTDLEKSHTKLOEIPVLHHEEQELLPNV---KEVSE---TQETMNELELTQESTTKD 1319
36 FILLINKLFSIDNN-----STEANTESL-----LKYIPEELNYSVEQKAGIEGVESR 85
1320 STTLARIEMERLRLNEKFQESQEIKSLTKERDNLKTKKEALEVHKDQJLK----EHRET 1375
86 VDLILFENDKKKASFNKKLKEAKKNEPIEDILIAEVKRPFTSFAKD----- 136
1376 LAKIQESQKQESQSLNMK-----EKDNETTK-----IVSEMEQ-----FKPKDSALLRIET 1422
137 -----KLKESDQLYRYLNOYQKHGILSNGKWLRYDKSKVLYCEKRYIBFNFKKIE 190
1423 MGLSKRLQESHDEMKSAKEDD-----LQRLQEVLOSQDLKENIKEIVA 1470
191 KEEYKEQEWFLFYILIRKERYLTKSNVIEVEKEIAKEKEIIQKTLKEILYERPDSDIV 250
1471 KHLTEEEELKVAHCCLKEQE-----ETINELRVNLSEKETEISTIQKLEAI-----NDKIQ 1522

QY 1035 HQNSRFFE-----DKDAKESKLLWIDKEDLEKVLTKDNQYQTERVFRATA-SNTNE 1087
 Db 1441 LQDEKRLERARIAQLEEELEEEQGNMEASDRVRKATQQAQBSNELATERSTAQKNESA 1500
 QY 1088 RTWISTLSPGNCYCVNSIYINDEKTPISLYKFLP-----IISIFNSFVDFLLRRFVDSN 1142
 Db 1501 RQQL-----BRQNKELRSKLHEMEGAVKSKFKSTI-----AALEAK 1536
 QY 1143 VLKSLYQCPMPQPEBEKEILSNPLYNLAKNTSLLIAKNDPENFKYLLYLEYFKPKDKV 1202
 Db 1537 IAQ--LEEQVEQEAQKATKSLKQDKKLEILLQVEDERKM-----AEQYKEQAERG 1589
 QY 1203 NKILKLDKEDPEKPEKENENNFIASLYSLAKE-DFIT-----LLG-DFKALKNNKKG 1253
 Db 1590 NARVKQLKQ--LEEAEEESQRINANRRKLQRLDEATESNEAMGREVNAKSKLRG 1644

RESULT 6
 US-10-995-561-661
 ; Sequence 661, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561
 ; NUMBER OF SEQ ID NOS: 85702
 ; SOFTWARE: Fast-Seq for Windows Version 4.0
 ; SEQ ID NO 661
 ; LENGTH: 1938
 ; TYPE: PRN
 ; ORGANISM: Homo sapiens
 US-10-995-561-661

Query Match 3.48; Score 224; DB 6; Length 1938;
 Best Local Similarity 19.9; Pred. No. 0.00024;
 Matches 274; Conservative 211; Mismatches 544; Indels 348; Gaps 60;

QY 61 KYIFEELNYSVEQOKA-----GQTEGVES-----RVDILLFENDKOKASFNNNKLEA 107
 Db 718 RIVFOEPRQRYEILANAIPKGMWDGQACILMIKALELDPNLYRICQSKIFRTGVLAH 777
 QY 108 KONEPIEDILIIAEVKRPTSFDAKDKLKSSEDQ--YRYLNQYQKHGYILSNKQW 165
 Db 778 LEEERDLKITDVIMAFQAM--CRGYLARKAFKRRQQQLTAMKVIQRNCAAYLKLNRNQW 835
 QY 166 RLVDKSKVLYGKRYIEFNFKTEEBEYKEQEWFLFYILIRKERYLKTSN--VIEVE 222
 Db 836 RLFTVKRPL-----LQVTRQEEWQAKED-----LQTKERQOKAENELKELEQK 881
 QY 223 KEQIAKEKEIIQKTL--KEILYERPDSDSIVFKIAKN-----IYDKFKL----- 264
 Db 882 HSQLTEKNLLQEQLOAETELAYAEAEVRVLAQKQELSELHEARLEEEEDRCQQL 941
 QY 265 --SGKEITQHILASILEESIFILRIFFIAYIEDNDIFKKILQBNKLYRSS-----ISFRY 318
 Db 942 QAERKKAQWQMLD--LEEQ-----LEEEAARQKQLEKVTAEAKIKLEDEI 987
 QY 319 FFYDENTKKKLEYKII-----TIFNLLDKGSDAIFPVPNGGLFSQDKV---YLNNE 369
 Db 988 LVMDQNNKLSKSKRLKEERISDITTTLAEE-----EKKAKNLTCLKNK 1031
 QY 370 GLLSISIEELVQMLFEEKNIKD-EKFKVYSRLDPKSKFGLVETLLDYDLRIADTVH 428
 Db 1032 HESMISLEVRLLK-----EKSQRELEKRLKLEGDASDFE-----QIADLQA- 1076
 QY 429 RIIEDGVYLIRTEBELE-----NKKVNKIATYKGNLYTSRLDRKSG-- 473
 Db 1077 QIABLMQAKBEELQAALRLDDEIAQKNNAKIKIRE-LEGHISDLQEDLDSEARA 1135

QY 474 -----AYTTPDDLTPFWJISSIEBQKTKSPLDIKIIDNSCGSHFLISCL 519
 Db 1136 KAEKOKRDLGEELEAKT--ELEDTLDSATQQLRAKREQEVTVL----- 1179
 QY 520 DYLTEKVMYELDFEDVKELDERVYRIESEEYDVQDSISK-----ELVRLMLLKK 573
 Db 1180 -----KKALDEETR-----SHEAQVQEMRQKHAQAVEELTEQLEQFQFR 1217
 QY 574 CIYGVNDINPISVEITWLSLWINTFFIGTPLSFIEHHIKAGNALL-----GYTKDE 623
 Db 1218 AKANLDKNKQTLEKNADLAGELRVLGQAKQVEHKKKLEAQVQELQSKCSGGERARAE 1277
 QY 624 FFDIVK--KPESGFSFLPKRIKEIITILEDIYQIKGINDT-----TKEDIKSKK 673
 Db 1278 LNDKVKHLQNEVESVTGMLNEAGKAIKADVASLSSQLQDTQELLQETRQKLVNYSK 1337
 QY 674 IYKEYEESKDI-----DNLRITPSLKLKLSYSPDKSLNMEFSDIASVISIENI 722
 Db 1338 LRQLEERNLQDLDEEMEAQNLERHISTLNI-QLSDSKK--KLQDFASTVEALEE- 1392
 QY 723 LGNKTSSSEDEKI-----EKIRKLSYYKFFHYGIEFPDIOEGFDIV----- 764
 Db 1393 -GKGRFQKEIENLTQOYEKAAAYDKLEKTKNRLQ-----QELDDLVDLDNOR 1440
 QY 765 --IGNPPMEKTKFNETEPFSKHPNVRKLGIKOEONLIKOEILSKDNHPLSI-----EYNE 817
 Db 1441 QLVSNLEKQKRPDQLLAEEKNISKY---ADERDRAEAREKETKALSLARALEALE 1497
 QY 818 EKNSIIAINNIYKFPDKFTSGDPNLFYFVTFNL-KLIKBNL--TYLVPSAIWNES 875
 Db 1498 AXEELERTNKLKAEWEDLVSSKDD-----VGKNVHELEKSKRALEQTQMEEMKTQLEEL 1551
 QY 876 SRILRKHIFARYKLVYI-----QFENKRFKDVHSSPKFAIFOLSNIKESISSPKAKFM 930
 Db 1552 EDELOATEDAKLRLEVNMQALKGQFERDLQARDEQNEEK-----RRQRLQRLHEYETELE 1606
 QY 931 IOSSDNIL-----KEITRDLKD--SKDDAYKGIELNINOIKKL-SPIOESITEFKD--- 978
 Db 1607 DERKQALAAAKKLEGLDKOLEQADSAIKGREAIKQLRKLQAKMDQFQLEEDARA 1666
 QY 979 --NEFTLINKFSPSALGEGYIDPKGLDPSIKNRK--SLLKECNKNLIFYSGANI 1034
 Db 1667 SRDEIFATAKENEKKAKSLEADLMQLEDLAAARAKQADLEKELEAELASSLGRNA 1726
 QY 1035 HQNSRFFE-----DKDAKESKLLWIDKEDLEKVLTKDNQYQTERVFRATA-SNTNE 1087
 Db 1727 LQDEKRLERARIAQLEEELEEEQGNMEASDRVRKATQQAQBSNELATERSTAQKNESA 1786
 QY 1088 RTWISTLSPGNCYCVNSIYINDEKTPISLYKFLP-----IISIFNSFVDFLLRRFVDSN 1142
 Db 1787 RQQL-----ERQNKELRSKLHEMEGAVKSKFKSTI-----AALEAK 1822
 QY 1143 VLKSLYQCPMPQPEBEKEILSNPLYNLAKNTSLLIAKNDPENFKYLLYLEYFKPKDKV 1202
 Db 1823 IAQ--LEEQVEQEAQKATKSLKQDKKLEILLQVEDERKM-----AEQYKEQAERG 1875
 QY 1203 NKILKLDKEDPEKPEKENENNFIASLYSLAKE-DFIT-----LLG-DFKALKNNKKG 1253
 Db 1876 NARVKQLKQ--LEEAEEESQRINANRRKLQRLDEATESNEAMGREVNAKSKLRG 1930

RESULT 7
 US-10-995-561-662
 ; Sequence 662, Application US/10995561
 ; Publication No. US20050272054A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 ; TITLE OF INVENTION: DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001559
 ; CURRENT APPLICATION NUMBER: US/10/995,561

QY 11 NPWISLVKQSKD-----FIKKENINKL-----KDF 36
| : : : : :
Db 314 NININIKKFKDKYFVEDSEKYSIDVESPKLYKSLMFGFTETNIAENYKIKTRASY 373
| : : : : :
QY 37 F-----ILFKNKL-----FSDINSTANI-----BSLKYPBELN-----YS 70
| : : : : :
Db 374 FSDSLPVKIKNLLDNEIYITIEBGFNISDKDMEKEYRGQKAKNQAYEISEHSLAVYK 433
| : : : : :
QY 71 VEOQKAGQIBGVSRVD--ILLFENDK---DKASFNNKLKEAKKN---EPIPIEDIL- 120
| : : : : :
Db 434 IOMCKSVKAPGICIDVDNEDLFIADKNFSDDLKSNRIEYNTQSGNIENDFPINELIL 493
| : : : : :
QY 121 ---IAEVKPT-----FSPD-----AKDKLKESEDOLVYLNQYQKHGKI-LS 160
| : : : : :
Db 494 DTLILSKIELPSENTSLTDFNDVDVYVEKQPAIKKIFTDENTIFQL--YSQTFPDIR 551
| : : : : :
QY 161 NGKVRLYDKSVLYGKRYIEFNFXKIEEYKEQWFLVFIYLRKERYLKTNSVIE 220
| : : : : :
Db 552 DISLTSSFDDA--LLFSNKVYSFSDMDYIKTANKVVEAGLPAGVW-----KQVNDVFIE 604
| : : : : :
QY 221 VEKQTAKEKIIQKTLKEILYERPDSDIVP-----KIANIYDKFKPLSGKEI 269
| : : : : :
Db 605 ANK-----SNTMDKI-----ADSLIVPYIGLALNVGNETAKGNFENAEFIAGAS- 649
| : : : : :
QY 270 TQHLASILEESIIFILRIFFI-AYIED-NDIEFKKILQENKL-----Y 310
| : : : : :
Db 650 ---ILLEFPELLIPVVGAFLESYIDNKNKIIKTI--DNALTKRNEKMSDYGVLIAQW 704
| : : : : :
QY 311 RSSISFRYPFYDENTKKLEY-----KKII-----TIFNLDDKGSDAIKFPVFNGLFSED 361
| : : : : :
Db 705 LSTVNTQFYTIKSGMYKALYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKL--- 760
| : : : : :
QY 362 KVKYLNNEGL-LSISEIEELVKMLPFEEKNIIDKFPVKYSRLDPKSGFELYETLLEYDL 420
| : : : : :
Db 761 -----NEGNOAIDNNPI-----NGCSVYLMKMKMIPLAVERKLLDFON 800
| : : : : :
QY 421 RIADTTVHRIEDGVYLRTEEBELNKKVNKIATYLGKNIYLSRSLDRKKSQAVYTPDD 480
| : : : : :
Db 801 TLKNLLNYIDENKLYLIGS-AEYKSKVN---YLK-----TIMPFDL-----SITYNDT 847
| : : : : :
QY 481 LTDFMVISSIEQLKTSPLDIKIIDNSCGSHFLISCLDYLTKVWYELDKPEDVKKEL 540
| : : : : :
Db 848 IL-LEMPKNYSILANNIILNRYKDN-----LIDLSCY-----CAKVEYVDGV--EL 893
| : : : : :
QY 541 DEEYRVIIIESEYVDQDSISKBLVLKMLLKXICVGVGINPISVEITMLSLMINTFI 600
| : : : : :
Db 894 NDKNQFKLTSSANSKIRVTQONIIFNSVFLDFSV-----SFWIRIPKY- 937
| : : : : :
QY 601 TPLSFIEHHIKAGNALLGYTKDFFDIVKKKFSGFSLFKKRIKEITILEDIYQKIKGI 660
| : : : : :
Db 938 -----KNDGQNYIHNETIINCWNNSGMKISIRGNRIIWLID-----ING- 980
| : : : : :
QY 661 NDTTKEDIEKSKIKIYEESKDIDNI--RIIFSLI-----KLYSLSPDKSLNMEFSD 711
| : : : : :
Db 981 -----NTKSVFPEYNIREDISEYINRWFFVTITNLLNNAKIY-INGKLESNTDIKD 1030
| : : : : :
QY 712 IASVISLIE-----NILGNKTSSEDKKEIKIRKLSYYSYKFFHYG 751
| : : : : :
Db 1031 IREVIANGELIIFKLDGDIRTOFTWVKYFSIFNTLSQSNIEBRYKIQSYSEYKDF-WG 1089
| : : : : :
QY 752 IEPFDIOEGFDIVGN-----PPWE---KTKFNETERFSKHIPNRYKLGIKBQN 797
| : : : : :
Db 1090 NPLMYNKETYFWNAGNKNVYKUKDSPVGEILLTRSKYNQ-----SKYI-NYRDLYIGERF 1145
| : : : : :
QY 798 IIKQ-----EILSKDNH 809
| : : : : :
Db 1146 IIRKNSQSINDDIVKEDY 1166
| : : : : :

RESULT 9

US-10-995-561-660

; Sequence 660, Application US/10995561

; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 660
; LENGTH: 1954
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-995-561-660

Query Match 3.4%; Score 219.5; DB 6; Length 1954;
Best Local Similarity 19.9%; Pred. No. 0.00041;
Matches 274; Conservative 213; Mismatches 544; Indels 349; Gaps 61;

QY 61 KVIPELNVSVQOQKA-----QOIEGVES-----RVDILLFENDKDKASPNKLEA 107
| : : : : :
Db 718 RIVQOEFQRORYEILAAINAIPKGMGDKQACILMIKALELDPNLYRIGOSKIPFRTGVLAH 777
| : : : : :
QY 108 KKNNEPIPIEDILIIAEVKRPTFSFDADKDLKESDQL--YRYLNQYQKHGILSGKW 165
| : : : : :
Db 778 LBEERDLKITVIMAFQAM--CRGYLARKAFAKRQOQLTAMKVIQRNCAAYLKLWNQW 835
| : : : : :
QY 166 RLYDKSVLYGKRYIEFNFXKIEEYKEQWFLVFIYLRKERYLKTNS---VIE 222
| : : : : :
Db 836 RLFTVKVPL-----LQVTRQEEEMQAKEDS-----LQTKERQQAENELKELEQ 881
| : : : : :
QY 223 KEQIAKEKEIIQKTL--KEILYERPDSDSIVFKIAKN-----IYDKFKL- 264
| : : : : :
Db 882 HSQLTEEKNLQEQQAETELYAEAEEMVRLEAAKQOELEELIHEARLEEBEEDRGOQL 941
| : : : : :
QY 265 --SGKEITOHILASILEESIIFILRIFFIAYIEDNDIFPKILQENKLYRSS---ISFRY 318
| : : : : :
Db 942 QAEKRWAAQOMLD--LEEQ-----LEEEAARQKQLEKVTABAKIKKLEDEI 987
| : : : : :
QY 319 FFYDENTKKKLYKII-----TIFNLDDKGSDAIKFPVFNGLFSEDVKV---YLNNE 369
| : : : : :
Db 988 LWMDQNNKLKSKERKLEBERISDLITNLAE-----EEKAKNLTKLKNK 1031
| : : : : :
QY 370 GLLTSISEIEELVKMLPFEEKNIKD-EKPVKYSRLDPKSGFELYETLLEYDLRIADTTVH 428
| : : : : :
Db 1032 HESMISELVRLKK---BEKSRQLEKLRKLEGDASDFHE-----QIADLQA- 1076
| : : : : :
QY 429 RIIEEDGVYLRTEEBELN-----NKKVNKIATYLGKNIYLSRSLDRKKS- 473
| : : : : :
Db 1077 QIAELKQWLAKKEEELQAAALRLDDEIAQNNALKKIRE-LEGHISDLQEDLDSEARA 1135
| : : : : :
QY 474 -----AYYTPDDLTDFMVISSIEBQLKTSPLDIKIIDNSCGSHFLISCL 519
| : : : : :
Db 1136 KAEKQKRDIGEELEALKT--ELEDTLDDSTATQELRAKREQEVTVL----- 1179
| : : : : :
QY 520 DYLTEKWYELDKFEDVKELDEEYRVIIIESEYVDQDSISK-----ELVLKRMLLKX 573
| : : : : :
Db 1180 -----KKALDEETR-----SHEAQVQEMRQKHAQAVEELTEQLRQPKR 1217
| : : : : :
QY 574 CIYGVDDINPISVEITMLSLMINTFIIFGTPLSFIEHHIKAGNALL-----GYTKDE 623
| : : : : :
Db 1218 AKANLDKNQKILEKENADLAGELRVLGQAQKQVEHHKKKKLEAQVQLQSKCSGERARAE 1277
| : : : : :
QY 624 FFDIVKK---KPFESGFSLPFKRIKEITILEDIYQKIKGINDT-----TKEDIEKSK 673
| : : : : :
Db 1278 LNDKVHKLQNEVESVTGMLNEAEGKAIKLAKOVASLSQLODTQELLQEBETROKLVNSTK 1337
| : : : : :
QY 674 IYKYEESKDI-----DNLRIIFSLIKLYSLSPDKSLNMEFSDIASVISLIENI 722
| : : : : :
Db 1338 LRQLEEBERNLQDQDDEEMAKONTERHISTLNI-QLSDSKK---KLQDPASTVEALEE- 1392
| : : : : :

```

QY 723 LGNTSSDEKKEI-----EKIRKLSYYKFFHYGIBFPDIOEGFDIV-----764
DB 1393 -GKRFQKEIENLTQOYBEKAAAYDKLETKRNLQ-----QELDDLVDLDNR 1440
QY 765 --IGNPWEKTFNTEFFSKHPIPNYKLGIGKIBONIIKOEILSKDNHPLSI-----EYNE 817
DB 1441 QLVSNLEKKQKQKFDQLLAEBEKNISKY---ADERDRAEAAREKETKALSARALEALE 1497
QY 818 EKNSIITAINNIVKDFPKCTSGDPNLFYVFTFNL-KLIKEKGNL-TYLVPSAINNESS 875
DB 1498 AKEELERTNMLKAEMEDLVSSKDD-----VGNVHELEKSKRALETQMEEMKTQLEEL 1551
QY 876 SRLRKHIIPARYKLANIY-----OPENKKRFDVHSSFKFAIFOLSNIKESTSSPKAKFM 930
DB 1552 EDELQATEDAKURLFVNMQALKGQFERDQARDEQNEEK-----RRQLQRLHEYETELE 1606
QY 931 IOSSDNIL-----KEITRDLK-----SKDDAYKGIELNINOIKKL-SPIOESIIEFKD---978
DB 1607 DERKQALAAAKKLEGLDKOLEQADSIAIKGREBAIKQLRKLQAKMDQFOLEEDARA 1666
QY 979 --NEEFTLINKMFSKFSALGEGYIDFKGLDPSIKNRK--SLLKECNKNLIFLYSGANI 1034
DB 1667 SRDEIPATAKENEKAKSLEADLMQLQEDLAAARARKQADLEKEBELABELASSLSGRNA 1726
QY 1035 HQNSRFPF-----DKDAKESKLLWIDKEDLVKLTQDNQYQTERVFRATA-SNTNE 1087
DB 1727 LODEKRLERLARIAQLEEELEEEQNGMEASDRVRKATQAQAEQLSNELATERSTAQKNSA 1786
QY 1088 RTWISTLSPGNCYCVNSIYINDEKTPISLYKFLF-----LISFNSFVDFLLRRFVDSN 1142
DB 1787 RQOL-----ERQNKELRSKLHEMEGAVKFKSTI-----AALEAK 1822
QY 1143 VLKSCILYQCPMPQPEEKEILSNPLYLNLAKNISLLIAKNDPENFKYLLYLEYFKDKVK 1202
DB 1823 IAQ--LEEVEQAREKQAATKSLKQDKKLEILAQVEDERKM-----AEQYKEQAEKG 1875
QY 1203 NKILKLDKDEFPYKVENENNFIASLYSLAKB-DFIT-----LLG-DPKALKNK-KKGED 1255
DB 1876 NARVKQLKQK--LEEAEBESQRIANRRKLQRELBATEESNEAMGVEVNAKSLKLRGNE 1933

RESULT 10
US-10-995-561-664
; Sequence 664, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 664
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-664

Query Match 3.4%; Score 219.5; DB 6; Length 1972;
Best Local Similarity 19.9%; Pred. No. 0.00042;
Matches 274; Conservative 213; Mismatches 544; Indels 349; Gaps 61;

QY 61 KYIFEELNYSVEQOKA-----GOIEGVES-----RVDILLFENDKOKASFNKKLEA 107
DB 718 RIVFQFRQRYEILAAAIKPGFMDGQACILMIKALELDPNLYRIGQSKIFPRTGVLAH 777
QY 108 KONEPIPIEDILIIAIVKRPPTSPDAKDKLKESEDOL--YRYLNQYQKHGYILSKNGKW 165
DB 778 LEEERDLKITDVIMAFQAM--CRGYLARKAFAKRQOOLTAMKYIQRNCAAYLKLNRQWQW 835

```

```

QY 166 RLYDKSKVLGYGKRYIEFNFXKIEBEKYEKEQEWFLFYILIRKERYLKTSN---VIEVE 222
DB 836 RLFTVKFPL-----LQVTRQEEEMQAKED-----LOKTKERQQAENELKELEOK 881
QY 223 KEQIAKEKEIIQKTL--KEILYERPDSDIVFKIAKN-----IYDKEFKL-----264
DB 882 HSQLTKEKNLLQEQLOAETELYABAEEMRVRLAAKQOELEELHEWEARLEBEEDRGQQL 941
QY 265 --SGKEITQHILASILEBSIIPILRIFFIAYIEDNDIPKILQENKLYRSS-----ISFVY 318
DB 942 QAERKMAQOMLD--LEEQ-----LEEBEAAARQQLQLEKVTAAEAKIKKLEDBI 987
QY 319 PFDYDENTKCKLYKKII-----TIFNLLDKGSDAKPVPVNGFLFSEDKVK--YLNNE 369
DB 988 LVMDONNKLKSKERLEERISDLTTLNLAEB-----BEKANMLTKLRNK 1031
QY 370 GLLSISEIEEILVLMFFEEKNIKD-EKPVKYSRLDPKSFGBELYETLLEYDRIADTTVH 428
DB 1032 HESMISELEVLKK---EKSROELEKLRKLEGDASDFHE-----QIADLOA- 1076
QY 429 RIIEGCVYLIRTEBELE-----NKKVNKIATYLGKNIYLTSRSLDRKKS--473
DB 1077 QJAEKLMQWLAKKEEELQALARLDDEIAQKNNAKKIRE-LEGHISDLQEDLDSERAARN 1135
QY 474 -----AYYTPDDLTDPMVVISSIEEQLTKSPDIKIIDNSCGSGHFLISCL 519
DB 1136 KAEQKQKRDIGEELEAKT--ELEDLTDSTATQELRAKEEQEVTVL-----1179
QY 520 DYLTEKVVYELDKFEDVKKELDEEYRVIIIEBSEEDVDQDSISK-----ELVKRMLLKK 573
DB 1180 -----KKALDEETR-----SHEAQVQEMRQKHQAQVLELTOEQFQR 1217
QY 574 CIYGVDIINPISVEITMLSIWINTPIFGTPLSPIEHKIKAGNALL-----GYTKDE 623
DB 1218 AKANLDKNQKTKLEKNADLAGELRVLQAKQVBEHKKKLEAQVQESLQSCSDGERARAE 1277
QY 624 FFDIVKK--KPESEGSFLPKKRIKIIITILEDIYQKIKGINT-----TKEDIKSKK 673
DB 1278 LNDKRVHKLQNEVSTVGMLENEAGKAIKLAQDVASLSQLODQTOELLQEBETQKLVSTK 1337
QY 674 IYKVEESKDI-----DNRIIFSLIKLYSLSPDKSLNMEFSDIASVISIENI 722
DB 1338 LRQLEEEERNSLQDQDEMEAKQNLERHISTANI-QLSDSKK---KLQDFASTVEALER- 1392
QY 723 LGNTTSSDEKKEI-----EKIRKLSYYKFFHYGIBFPDIOEGFDIV-----764
DB 1393 -GKRFQKEIENLTQOYBEKAAAYDKLETKRNLQ-----QELDDLVDLDNR 1440
QY 765 --IGNPWEKTFNTEFFSKHPIPNYKLGIGKIBONIIKOEILSKDNHPLSI-----EYNE 817
DB 1441 QLVSNLEKKQKQKFDQLLAEBEKNISKY---ADERDRAEAAREKETKALSARALEALE 1497
QY 818 EKNSIITAINNIVKDFPKCTSGDPNLFYVFTFNL-KLIKEKGNL-TYLVPSAINNESS 875
DB 1498 AKEELERTNMLKAEMEDLVSSKDD-----VGNVHELEKSKRALETQMEEMKTQLEEL 1551
QY 876 SRLRKHIIPARYKLANIY-----OPENKKRFDVHSSFKFAIFOLSNIKESTSSPKAKFM 930
DB 1552 EDELQATEDAKURLFVNMQALKGQFERDQARDEQNEEK-----RRQLQRLHEYETELE 1606
QY 931 IOSSDNIL-----KEITRDLK-----SKDDAYKGIELNINOIKKL-SPIOESIIEFKD---978
DB 1607 DERKQALAAAKKLEGLDKOLEQADSIAIKGREBAIKQLRKLQAKMDQFOLEEDARA 1666
QY 979 --NEEFTLINKMFSKFSALGEGYIDFKGLDPSIKNRK--SLLKECNKNLIFLYSGANI 1034
DB 1667 SRDEIPATAKENEKAKSLEADLMQLQEDLAAARARKQADLEKEBELABELASSLSGRNA 1726
QY 1035 HQNSRFPF-----DKDAKESKLLWIDKEDLVKLTQDNQYQTERVFRATA-SNTNE 1087
DB 1727 LODEKRLERLARIAQLEEELEEEQNGMEASDRVRKATQAQAEQLSNELATERSTAQKNSA 1786
QY 1088 RTWISTLSPGNCYCVNSIYINDEKTPISLYKFLF-----LISFNSFVDFLLRRFVDSN 1142

```

Db 1787 RQQL-----ERQNELRSKLHEMGA VKSKFKSTI-----AALEAK 1822
QY 1143 VLKSCLYQCPMPQPEBEKEILSNPLYNLAKNTSLIAKNDPENFKYLLYLEVFKPKYK 1202
Db 1823 IAQ--LEEQVEQBEAREQAATSLKQKDKKLEILLQVEDERKM-----AEQYKEQAEKG 1875
QY 1203 NKLKLDKDEPFKEKENENNFIIASLYSLAKE-DFIT-----LLG-DFKALKNK-KKGED 1255
Db 1876 NARVKQLKQK--LEEAEEBSQRINANRRKLQRELDATESNEAMGREVNALSKLARGNE 1933

RESULT 11
US-10-995-561-666
; Sequence 666, Application US/10995561
; Publication No. US200502720541
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-666

Query Match 3.4%; Score 219.5; DB 6; Length 1972;
Best Local Similarity 19.9%; Pred. No. 0.00042;
Matches 274; Conservative 213; Mismatches 544; Indels 349; Gaps 61;

QY 61 KYIFEELNYSVEQOKA-----GOIEGVES-----RVDILLFENDKOKASPNKKLEA 107
Db 718 RIVQEFQRYEYILANAIPKGMWGQACILMIKALELDPNLYRIGOSKIFPRTGVLAH 777
QY 108 KQNEPIPIEDILIIAEVRKPTSPDAKDKLKESEDL--YRYLNOYQKHGYLSNGKVM 165
Db 778 LEEERDLKITDVIMAFQAM--CRGYLARKAFARQQQLTAMKYIQRNCAAYLKLNRWQW 835
QY 166 RLYDKSKVLYGKRYIEFNFKLKEEYKQEWFLFYILIRKERYLTSN---VIEVE 222
Db 836 RLFTKVKPL-----LQVTROEEEMQAKED-----LQTKERQQAENELKELEQ 881
QY 223 KEQIAKEKEIIQKTL--KEILYERPDSDIVFKIAKN-----IYDKFEKL----- 264
Db 882 HSQITBEKNLLQEQLOAETELYAABEMRVLAACKQOELEELIHEMARLEEEEDRGOOL 941
QY 265 --SKGEITQHILASILEESIIFILRIPPIAYIEDNDIFKKILQENKLYRSS-----ISPRY 318
Db 942 QAERKMAQOMLD--LEEQ-----LEEBAARQKLQLEKVTAETAKIKLEDEI 987
QY 319 PFYDENTKKKLEVKKII-----TIFNLLDKGSDAIKFPVNGSLPSEDKVK---YINNE 369
Db 988 LVNDDQNNKLKSKERKLEERISDLTTNLAE-----EEKAKNLTKLK 1031
QY 370 GLLSISIEBILVKMLPFPEKNKID--EKFKYSRLDPKSPGELVETLLEVDLRIADTTVH 428
Db 1032 HESWISLEVLK-----EEKSQEKLKRRKLEGDASDFE-----QIADLOA- 1076
QY 429 RIIEDGVYLIRTEEBE-----NKKVNKIATYKGNIVLTSRLDRKSG-- 473
Db 1077 QIAELKQMLAKEEELQAALARDLDETAQKNALKKIRE--LEGHISDLQDLDSERAARN 1135
QY 474 -----AYTTPDLDTFMWVIEISIEBOLKTSPLDIKIIDNSCCSGHFLISCL 519
Db 1136 KAEKQKRDLEGELEALKT--ELEDTLDSTATQQLRAKREQEVTVL----- 1179
QY 520 DYLTEKVMYELDKPVEDVKELDEYRVIIESEBEYVDQDSISK-----ELVLKRMLLKX 573

Db 1180 -----KKALDEETR-----SHEAQVQEMRQKHAQVAEELTEQLQPFQKR 1217
QY 574 CIYGVNDINPISVEITWLSIWNITFFCTPLSFIEHHIKAGNALL-----GYTKDE 623
Db 1218 AKANLDKNNKTLEKENADLAGELRVLGQAQVEHKKKKLEAQVQELQSKCSGDERARAE 1277
QY 624 PFDIVKK--KPESGFSJFKKRIEIIITLEDIYQKIGINDT-----TKEDIEKSKK 673
Db 1278 LNDKVHKLQNEVESVTGMLNEAGKAIKLAKDVASLSQLODTQELLQEBETQKLNVSFK 1337
QY 674 IYKYEESKDI-----DNLRIIFSLIKYLSLSPDKSLNMEFSDIASVISLIENI 722
Db 1338 LQLEERNSLQDLDEEMEAQNLERHISTLNI--QLSDSKK--KLQDFASTVEALEE- 1392
QY 723 LGNTSSSEDKKI-----EKIRKLSYYFFHYGIEFFDIOEGFDIV----- 764
Db 1393 -GKGRFOKEIENLTQYBEKAAAYDKLEKTKNRLQ-----QELDDLVDLDNOR 1440
QY 765 --IGNPPMEKTFNTEFPFSKHIPYVRKLGIKOEONIIKOEILSKDNHPLSI-----EYNE 817
Db 1441 QUVSNLEKKQRFQDQLAEBEKNISSKY--ADDERAEEAREKETKALSARALEEAL 1497
QY 818 EKNSIIAINNIYKDFKCTSGDPNLFYFVTNL-KLIKEGML--TYLVPSAIWNES 875
Db 1498 AKEELERTNKMKAEMEDLVSSKDD-----VGKVVHELEKSKRALETQMEEMKTQLEEL 1551
QY 876 SRILRKHFARYKLVNIY-----QFENKGRFKDVHSSPKFAIFQLSNIKESTSSFKAKPM 930
Db 1552 EDELQATEDAKLRLEVNMQALKQGFERDLQARDEQNEEK-----RRQLQRLHEYETELE 1606
QY 931 IQSSDNIL-----KEITRDLKD--SKDDAYKGIELNINOIKKL-SPIOESIIEPKD--- 978
Db 1607 DERKQALAAAKKKLEGLDKLELQADSAIKGREAIKQLKRLQOMKDFORELEDARA 1666
QY 979 --NEEFTLINMFKSPSALGEGYIDFKGLDPSIKNRK--SLKKECNKNKILFLYSGANI 1034
Db 1667 SRDEIFATAKENEKKAKSLEADLMQLQEDLAAERARKQADLEKEBELASSLSGRNA 1726
QY 1035 HGFNRRFFE-----DKAKESSKLLWIDKELEKVLTKDNQYQTVRYRAIA-SNTNE 1087
Db 1727 LQDEKRRLERLAQLEEELEEBEQNGMEAMSDRVKATQAQAEQLSNELATERSTAQKNESA 1786
QY 1088 RTMISTLSPGNCVNSIYINDEKTPISLYKKLF-----IISFNSFVDFLLRRPVDSN 1142
Db 1787 RQQL-----ERQNELRSKLHEMGA VKSKFKSTI-----AALEAK 1822
QY 1143 VLKSCLYQCPMPQPEBEKEILSNPLYNLAKNTSLIAKNDPENFKYLLYLEVFKPKYK 1202
Db 1823 IAQ--LEEQVEQBEAREQAATSLKQKDKKLEILLQVEDERKM-----AEQYKEQAEKG 1875
QY 1203 NKLKLDKDEPFKEKENENNFIIASLYSLAKE-DFIT-----LLG-DFKALKNK-KKGED 1255
Db 1876 NARVKQLKQK--LEEAEEBSQRINANRRKLQRELDATESNEAMGREVNALSKLARGNE 1933

RESULT 12
US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT

Db 226 VOTLMDNMTLERERESEKLECKLQSQVASEATISQLRSELAKPGQOEVAVYVQELQK 285
Qy 44 LFSIDDNSTAN---TESLLKYIFEELNYSVEQKAGQIEGVESRDVILLFENDKDKASF 100
Db 286 LKSSVNELTKQNTLTENLLK---KEQDYTKLEERINEESVSKNIQATLHOKDLDCQQL 342
Qy 101 NNKLKBAKKNNEPIETIILLIAEVRPTFSFPAKDKLJESDQL---YRYL----- 149
Db 343 QSFLSASETSLRHIVB---LSE-----KGEATQKLAEELSEVETKYOHLKABFKQLQ 392
Qy 150 --NQYQKHGYILNGKWRLYDKSVLYGKRYIEFNFXXKIBKBY-----KEQWPF 200
Db 393 QOREEKEOHLQ-QSEINQLH---SKLLETERQJGEAH-GRLEQRQLSSEKLMKDEQOVA 449
Qy 201 VLFIVILIRERYL---KTSNVIEKE-----QIAKEKEIIOKTLEILYERPDSDIVFKI 253
Db 450 DLQLKLSRLEQKEKVTNSTELQHOLDTKYQHQEQALQSQSTTAKUREAOND---LEQV 507
Qy 254 AKNIYKFKLSGKEITOHILASILBESIIFILRIFFIAYIEDNDIFKKI----- 303
Db 508 LRQIGDKQKIONLE---ALLQSKENISLLEK-----EREDLYAKIQAGEGETAVL 556
Qy 304 -----LQE-----NKLYRSSISFRYFFVDENNTKKLYKKIITIFNLLDKGSDAI 348
Db 557 NQLQEKNNHTIQEQVTLQTEKLKNQSESHKQ--AQENLHDQVQSKA-----HLRAAQDRV 609
Qy 349 KPPVFNGLPSEDKVXYLN---NEGLLSISEIEILVLMFLPFEKNIKDEKFKVYKSELDP 405
Db 610 -----LSLETSVNELNSQLESKEVSQL-DIQIK-----AKTELLS----- 646
Qy 406 KSPGELYETLLELDRIADTTHRIIEDGVYLIRTEELENNKKVNKIATVYKGNIVLTGR 465
Db 647 ---AEAATAQRADLQNHLDTAQNALQD-----KQELNKITITOLD---QVTAK 689
Qy 466 SLDRKXGAYTTPDDLTDFWVSSIEBQTKTSPLDIKIIDNSCGSHFLISCLDYLTREX 525
Db 690 LQDKQEH-----CSQLESHLKEYKEVLS-----LEQKTEE 720
Qy 526 VWTELQKFEVDKELDEYRVIIIESEEDVDQD-----SISKELVLRM 569
Db 721 LEGQIKKLEADSLE-----VKASKBQALQDLQOQORQLNTDLRLATLSKQLEMEKE 772
Qy 570 LIAKXCIYGVDPINPISVEITMLSWINTFIPGTPLSPTEHHIKAGNALLGYTKDEFEDIVK 629
Db 773 I-----VSTFLDL-----Q 782
Qy 630 KKFESGFLPKRKIKSIITILEDIYQIKGINDTTKEDIEKSKKIYKVEES----- 681
Db 783 KKSEA-----LESIKQKL-----TKQ--EEKKILKQDPETLSQETKIQ 819
Qy 682 -KDIDNLIIFSLIKLYSLSPDK-SLNWPSDIASVISLIENILGNKTSSEDEKIEKIR 739
Db 820 HEELNN-RIQTTVTELQVKMEKALMTSLTVKDLKSVSDSLKNSKSEFEKEN----- 873
Qy 740 KLSYKYFFHYGIEFPDIQSGPDVIGNPPWEKTKFNETFEFFSKHIPNYRKLGIEKQNI 799
Db 874 -----QKG-----KAAILDLEKTCELKH--QLQVQMENTL 902
Qy 800 KQE-----ILSKD---NHPLSIBYNEKNSIIANNIYKDFDKPTSGGPNLFRYFVTF 851
Db 903 KEQKELKSLSEKEASHQLKLELNSMOEQLIQAQNTLKQNEK-----EEQQLQ 952
Qy 852 NLKLIKEKGNLTIVLPSAINWESSRILRKHIFARYKLVNIYFENK----- 898
Db 953 NINELKQSSBQKKQKEALQGELKIAVLQK-----TELENKLOQLTQAAQELA 1001
Qy 899 -----KRFKDVHSSFPKPAIFOLSNIKESTSSFKAKFMIOSSDNI----- 937
Db 1002 AEKEKISVLQNNYKESQETFKQIQSDPYGRESILLATROLKSVKELSLAQSDLSNRN 1061
Qy 938 -----LKEITRDLKDSK-----DDAYKGIELNINQIKKLSPIQBSIIIEFKDNBSFTLINKM 988
Db 1062 QIGNQNKLIQELKTAKATLQDSAKKEQQQLQ-ERCKALQDIQKE-----KSLKEKELVNEK 1116

Qy 989 FSKSALGEGYIDFKK---GLDPSIKNRK-SLLKECNN---KNLIF-----L 1028
Db 1117 -SKLAEIEIKCRQKEITKLNELKSHKLESIKETITNLKDAKQALLIQOKLELQKADSL 1175
Qy 1029 YSGANIHQPNRPFDDKADAKESSKLLWIDKEDLEKVLTKDNQYQTVRVPYRAIASNTNR 1088
Db 1176 KAAVEQEKRNQOILKDQVKKEBEEL---KKEFIEKEAKLHSEIKKEVEGMKKHEENAKL 1232
Qy 1089 TMIISTLSPGNCYCVNSIYINDEKTPISLYKK 1119
Db 1233 TMOITALNENLGTVKKEWQSSQRRVSELEKQ 1263

RESULT 14

US-10-793-626-1660
; Sequence 1660, Application US/10793626
; Publication NO. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1660

Query Match 3.2%; Score 206.5; DB 6; Length 885;
Best Local Similarity 19.0%; Pred. No. 0.00075;
Matches 205; Conservative 188; Mismatches 313; Indels 371; Gaps 53;

Qy 17 YQLSKDFIKKENINKLKDFFILIKNKLFSIDNSTEANIESLLKYIFBELNYSVBQOKA 76
Db 14 YQLSKEMEQSDVIVTVSDI-----DHYTEDN-----QRLDERLNLHLSQ-- 54
Qy 77 QGIEGVESRDVILL-----FENDKDKASFNNKLKEAKNNEPIEDILIIAEVKRPT 129
Db 55 AEKEGQQAQINQLQKYKGKQOONDYDIEKLANVELVKATENYEQLS-GKLNVLLEERKQ 113
Qy 130 FSFDAK--DKLASEDOLYRNLNOKYKHGYILNGKWRLYDKSVLYGKRYIEFNFXX 187
Db 114 SETNARYEBELNLESQIDSIKNEKAQNEKLIA-----ELKNQKQKLNKEVQLESLLYI 168
Qy 188 IBEKEEYKEQWFLVIYILIRKERYLKTNSVIEVEKEQIAKEKEIITQTL--KEILYERP 245
Db 169 SDEQHDKELEE-----IKNSYITLMS-----EQSDVNDIRFLEHTINENEAKSRL 215
Qy 246 DSIV--FKIKNY-----DKFPLSKGK-----EITQHI--LASILEBSIIFILRFFIA 292
Db 216 DSRLEAFNQLDKIQONTQTOKEYOSKSKMEKVEQNIQQLFQQLTDS----- 264
Qy 293 YIEDNDIDFKKILQ--ENKLYRSSISFRYFFYDENTKKLEYKKIITIFNLLDKGSDAIKF 350
Db 265 -----KRLISEYENKLYQA-----YRYNEKLKSRID-----SLATQEEEDYTYF 302
Qy 351 PVFNGG--LFSDEK----- 362
Db 303 --FNGVKHILKAKDKLRGIGHGAVAEVINVPSEMTQAIETALGASLQHVIVDNEKDGROA 360
Qy 363 VKYLNNEGL-----LSISEIEILVQMLPFEKNIKD-----EKFKYSRLDPKSPGELY 412
Db 361 IQYLRKQGLGRATFLPLNVIPQPHV-----AADINDVARGSQGFINIAS-DAINVSAYK 413

413	QY	ETLLEVDLRIADTTVHR	IEDGVYLIRTEEBLEN	--KVUNKIATYLGKNIYLTSTRSLDRK	470	
414	Db	QMIIE---NLLGNTI--	I--VEN---	LKHANELABAIRITRIVT-LEGDVVNPFGSM--T	461	
471	QY	KSGAYYTPDDLDFW	VISSIEBQKTKSP	LDIKIIDNSCGSHFLISCLDYLPEKVVYEL	530	
462	Db	GGGARKTSILSQ	DELSTMRNQL	-----DYQRQTAEF-	496	
531	QY	DPEDVK----	KELDEYRVII	EESEY--DVQDSI-SKELVLKMLLKCCYIGVDINPISV	585	
497	Db	RQFKEQKTA	EQQLSEQY--	PSASQVYNLKEQVHHLELDRLKTO-----	540	
586	QY	EITMLSLWINTI	IFGTPSLPST	TEHHIKAGNALLGYTKDEFPD	IVKKKPFSGSLPKKRIKE	645
541	Db	-----	EAHLKNEHEFE	FEKNDGYQSDKS-----KETLKE	570	
646	QY	IITILEDIYQIK	INGINDTTKEDI	EKSKKIYKEYEESKDIDNLRIP	SLIKLYSLSPDKSL	705
571	Db	KQNHLEIOOQL	KQL-----	ESDIERYTQLSEKGAS-----THQTOOQL	610	
706	QY	NMEPFSIAS	VI	SLIEN-----ILGNKTS	SSDKIEKIRKLSYKFFHYGIEFPDIOE	759
611	Db	HQKQSDLA	VVKERIKSQ	KQVYERLDKQLSDSERQKIEVNEKI-----KLFN-	656	
760	QY	GFDIVIGNP	PWEKTK-----	PNE--TEPFSKHPNVRKLGIKQKNIUK--QEI	803	
657	Db	SDMMGKDA	FAFKLEQIQOQ	QOQENVQRNLOQLSEIKOQRKDLNEKIBINESQLQKHODI	715	
804	QY	LSKDNHPLS	IEYNEK----	NSIIANNIYKDFCFTSGGDPNLFR	YVTFVFNKLKLIK	857
716	Db	LSTENHYQDI	KAKQSKLDVL	INHAI	DHLNDTY-----QLTV	751
858	QY	EKGNTIYLP	SPAIWNESSRIL	KHII-----FARYKLN	IYIQPENKKRKFQDVHSSP	908
752	Db	ERARWEYD	SEIDTIDN-----	LRKKVKLT	KTKTIDELGPVNLNALEQEF-----ELNERY	799
909	QY	KPAIFOLSN	IKESTGSF	KAQFMIOGSDNLIK	EITRDLKSDKDAYKGIELNINQIKK	965
800	Db	TFINEORTDL	RRAK-----	ETLEOI	IHEMDKEVEGPKTTFHFAVOHFTTVFK	847

```

RESULT 15
US/11/062
; Sequence 4, Application US/11062471A
; Publication No. US20050255093A1
; GENERAL INFORMATION:
; APPLICANT: SHONE, Clifford Charles
; APPLICANT: SUTTON, John Mark
; APPLICANT: HALLIS, Bassam
; APPLICANT: SILMAN, Nigel
; TITLE OF INVENTION: Delivery of Superoxide Dismutase to Neuronal Cells
; FILE REFERENCE: 1581.08000001
; CURRENT APPLICATION NUMBER: US/11/062.471A
; CURRENT FILING DATE: 2005-02-22
; PRIOR APPLICATION NUMBER: 09/831,050
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: PCT/GB99/03699
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: GB 9824282.9
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Construct comprising Mn-SOD from B. stearothermophilus, a linker,
US/11/062.471A-4

```

Query Match 3.2%; Score 206.5; DB 7; Length 1070;
Best Local Similarity 20.3%; Pred. No. 0.00094;
Matches 202; Conservative 171; Mismatches 353; Indels 271; Gaps 51;

Search completed: January 24, 2006, 20:55:47
Job time : 30.4381 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 14.5258 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-18
Perfect score: 329
Sequence: 1 LAATPLVDDTVIGKLRITAKI.....GAIDPQTYHYKNEAIIEL 65

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329	100.0	65	8	ADU98759 Borrelia
2	329	100.0	160	8	ADU98803 Borrelia
3	329	100.0	370	8	ADU98805 Borrelia
4	329	100.0	370	8	ADU98761 Borrelia
5	73	22.2	233	5	AB55144 Lactococc
6	67	20.4	188	8	ADT59925 Plant pol
7	65	19.8	284	6	ABM73458 Staphyloc
8	64.5	19.6	348	7	ABM90374 Rice abio
9	64.5	19.6	601	6	ABU33885 Protein e
10	64	19.5	138	8	ADY05830 Plant ful
11	64	19.5	148	8	ADY58407 Plant pol
12	64	19.5	193	8	ADY04902 Plant ful
13	64	19.5	193	8	ADY78687 Plant ful
14	64	19.5	193	8	ADY22628 Plant ful
15	64	19.5	193	8	ADY4536 Plant ful
16	63.5	19.3	301	8	ADN17923 Bacterial
17	63	19.1	380	6	ABU29425 Protein e
18	63	19.1	381	7	ADH87344 Enterococ
19	63	19.1	491	2	AAW09771 Pichia ac
20	62.5	19.0	402	8	AD29219 Bacterial
21	62.5	19.0	404	7	ABM86837 Rice abio
22	62.5	19.0	445	9	ADY71564 Laccase p
23	62.5	19.0	1017	8	AD411736 Bacterial
24	62	18.8	351	9	ABM93745 M. xanthu

25	62	18.8	580	6	RAO19831.
26	62	18.8	750	6	AAE36322 Arabidops
27	61.5	18.7	122	6	ABU24382 Protein e
28	61.5	18.7	302	4	AAG82943 S. epider
29	61.5	18.7	302	6	ABU43297 Protein e
30	61.5	18.7	327	7	ABO64173 Klebsiell
31	61.5	18.7	350	5	ABP38538 Staphyloc
32	61.5	18.7	350	8	ADS06181 Staphyloc
33	61.5	18.7	695	4	ABG28835 Novel hum
34	61	18.5	355	4	AAU35146 Enterococ
35	61	18.5	355	6	ABU29086 Protein e
36	61	18.5	355	8	ADH97192 E. faecal
37	61	18.5	355	8	ADH97194 E. faecal
38	61	18.5	355	9	AEC13547 Enterococ
39	61	18.5	355	9	AEC13549 Enterococ
40	61	18.5	371	7	ADH87288 Enterococ
41	61	18.5	384	4	AAG92653 C glutami
42	61	18.5	388	4	AAAB79449 Corynebac
43	61	18.5	388	4	AAAB79450 Corynebac
44	60.5	18.4	645	7	ABO76507 Pseudomon
45	60	18.2	144	3	AA58826 Breast an

ALIGNMENTS

RESULT 1
ADU98759
ID ADU98759 standard; protein; 65 AA.
XX
AC ADU98759;
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 18.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
PN WO2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
XX (TEXA) UNIV TEXAS SYSTEM.
XX
PI Sykes KF, Hale KS, Johnston SA;
XX
DR WPI; 2004-834155/82.
XX
PT Immunizing a subject against Borrelia burgdorferi infection comprises providing to the subject at least one Borrelia antigen or its fragment.
XX
PS Claim 27; SEQ ID NO 18; 121pp; English.
XX
CC The invention describes a method of immunizing a subject comprising providing to the subject at least one Borrelia antigen or its fragment to induce an immune response. Also described are: an isolated polynucleotide comprising a sequence having at least 17 contiguous nucleotides in common with a sequence not given in the specification; an isolated polypeptide having at least 5 consecutive amino acids of the sequence not given in the specification; a vaccine composition comprising at least one Borrelia antigen or at least one polynucleotide encoding a Borrelia antigen; screening for at least one test polypeptide or test polynucleotide encoding a polypeptide for an ability to produce an immune response; preparing a vaccine; vaccinating a subject; treating a subject infected

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 65 AA;

Query Match 100.0%; Score 329; DB 8; Length 65;
 Best Local Similarity 100.0%; Pred. No. 1e-35;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKEGVDLAGAIDEQFTYHIKNE 60
 |||||
 DB 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKEGVDLAGAIDEQFTYHIKNE 60

QY 61 AIIEL 65
 |||||
 DB 61 AIIEL 65

RESULT 2
 ADU98803
 ID ADU98803 standard; protein; 160 AA.

XX AC ADU98803;
 XX DT 24-FEB-2005 (first entry)
 XX DE Borrelia burgdorferi antigenic polypeptide seqid 62.
 XX KW antibacterial; vaccine; immune stimulation; immunity; antigen;
 XX KW DNA library.

XX OS Borrelia burgdorferi.
 XX PN WO2004103269-A2.
 XX PD 02-DEC-2004.
 XX PF 17-OCT-2003; 2003WO-US033056.
 XX PR 18-OCT-2002; 2002US-0419401P.
 XX PA (MACR-) MACROGENICS INC.
 XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Sykes KF, Hale KS, Johnston SA;
 XX DR WPI; 2004-834155/82.
 XX DR N-PSDB; ADU98802.
 XX PT Immunizing a subject against Borrelia burgdorferi infection comprises
 XX PT providing to the subject at least one Borrelia antigen or its fragment.
 XX PS Claim 27; SEQ ID NO 62; 121pp; English.

CC The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be

CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 160 AA;

Query Match 100.0%; Score 329; DB 8; Length 160;
 Best Local Similarity 100.0%; Pred. No. 3.1e-35;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKEGVDLAGAIDEQFTYHIKNE 60
 |||||
 DB 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKEGVDLAGAIDEQFTYHIKNE 60

QY 61 AIIEL 65
 |||||
 DB 61 AIIEL 65

RESULT 3
 ADU98805
 ID ADU98805 standard; protein; 370 AA.

XX AC ADU98805;
 XX DT 24-FEB-2005 (first entry)
 XX DE Borrelia burgdorferi antigenic polypeptide seqid 64.
 XX KW antibacterial; vaccine; immune stimulation; immunity; antigen;
 XX KW DNA library.

XX OS Borrelia burgdorferi.
 XX PN WO2004103269-A2.
 XX PD 02-DEC-2004.
 XX PF 17-OCT-2003; 2003WO-US033056.
 XX PR 18-OCT-2002; 2002US-0419401P.
 XX PA (MACR-) MACROGENICS INC.
 XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Sykes KF, Hale KS, Johnston SA;
 XX DR WPI; 2004-834155/82.
 XX DR N-PSDB; ADU98804.
 XX PT Immunizing a subject against Borrelia burgdorferi infection comprises
 XX PT providing to the subject at least one Borrelia antigen or its fragment.
 XX PS Claim 27; SEQ ID NO 64; 121pp; English.

CC The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX Sequence 370 AA;

Query Match 100.0%; Score 329; DB 8; Length 370;
 Best Local Similarity 100.0%; Pred. No. 8.7e-35;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAKGVDLGAGGAIHQFTYHYIKNE 60
 |||||
 DB 211 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAKGVDLGAGGAIHQFTYHYIKNE 270
 |||||

QY 61 AIIEL 65
 |||||
 DB 271 AIIEL 275
 |||||

RESULT 4
 ADU98761
 ID ADU98761 standard; protein; 370 AA.

XX AC ADU98761;
 XX DT 24-FEB-2005 (first entry)
 XX DE Borrelia burgdorferi antigenic polypeptide seqid 20.
 XX KW antibacterial; vaccine; immune stimulation; immunity; antigen;
 XX KW DNA library.

XX OS Borrelia burgdorferi.

XX FN WO2004103269-A2.

XX PD 02-DEC-2004.

XX PF 17-OCT-2003; 2003WO-US033056.

XX PR 18-OCT-2002; 2002US-0419401P.

XX PA (MACR-) MACROGENICS INC.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Sykes KF, Hale KS, Johnston SA;

XX DR WPI; 2004-834155/82.

XX DR N-PEDB; ADU98760.

XX PT Immunizing a subject against Borrelia burgdorferi infection comprises
 providing to the subject at least one Borrelia antigen or its fragment.

XX PS Claim 27; SEQ ID NO 20; 121pp; English.

XX The invention describes a method of immunizing a subject comprising
 providing to the subject at least one Borrelia antigen or its fragment to
 induce an immune response. Also described are: an isolated polynucleotide
 comprising a sequence having at least 17 contiguous nucleotides in common
 with a sequence not given in the specification; an isolated polypeptide
 having at least 5 consecutive amino acids of the sequence not given in
 the specification; a vaccine composition comprising at least one Borrelia
 antigen or at least one polynucleotide encoding a Borrelia antigen;
 screening for at least one test polypeptide or test polynucleotide
 encoding a polypeptide for an ability to produce an immune response;
 preparing a vaccine; vaccinating a subject; treating a subject infected
 with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 ; and detecting antibodies to a herpesvirus. The method is useful in
 immunizing a subject against Borrelia burgdorferi infection. This is the
 amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 useful in immunizing a subject against infection by a member of the
 Borrelia species.

XX SQ Sequence 370 AA;

Query Match 100.0%; Score 329; DB 8; Length 370;
 Best Local Similarity 100.0%; Pred. No. 8.7e-35;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAKGVDLGAGGAIHQFTYHYIKNE 60
 |||||
 DB 211 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAKGVDLGAGGAIHQFTYHYIKNE 270
 |||||

QY 61 AIIEL 65
 |||||
 DB 271 AIIEL 275
 |||||

RESULT 5
 ABB55144
 ID ABB55144 standard; protein; 233 AA.

XX AC ABB55144;

XX DT 29-AUG-2003 (revised)

XX DT 16-MAY-2002 (first entry)

XX DE Lactococcus lactis protein ysfD.

XX KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX OS Lactococcus lactis; ILL403.

XX FN FR2807446-A1.

XX PD 12-OCT-2001.

XX PF 11-APR-2000; 2000FR-00004630.

XX PR 11-APR-2000; 2000FR-00004630.

XX PA (INRG) INRA INST NAT RECH AGRONOMIQUE.

XX PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX DR WPI; 2002-043418/06.

XX PT New nucleotide sequence useful in the identification of Lactococcus
 lactis and related species.

XX PS Claim 6; SEQ ID NO 1846; 2504pp; French.

XX The present invention is related to a Lactococcus lactis nucleotide
 sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
 acid sequence is useful in the detection and/or amplification of nucleic
 acid sequence, particularly to identify Lactococcus lactis or related
 species. The proteins of the invention are useful for the biosynthesis or
 biodegradation of a composition of interest. The invention helps research
 in lactic bacteria, particularly useful in the production of yogurt and
 cheese. Note: The sequence data for this patent is based on equivalent
 patent WO200177334 (published 18-OCT-2001) which is available in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX SQ Sequence 233 AA;

Query Match 22.2%; Score 73; DB 5; Length 233;
 Best Local Similarity 37.0%; Pred. No. 0.53;
 Matches 20; Conservative 7; Mismatches 17; Indels 10; Gaps 2;

QY 22 FYSLLNETGLD-----GVPAKGVDLGAGGAIHQFTYHYIKNEAIIEL 65
 |||||
 DB 101 FYNKLNIGADWARGITDFKGLGLPGTKDEAEQDFIYYSIRYEDIMSV 154
 |||||

RESULT 6
 ADT59925
 ID ADT59925 standard; protein; 188 AA.
 XX AC ADT59925;

XX DT 02-JUN-2005 (first entry)
 XX DE Rice abiotic stress responsive polypeptide SEQ ID NO:9096.
 XX KW abiotic stress tolerance; transgenic plant; plant; cereal; agriculture.
 XX OS *Oryza sativa*.
 XX PN WO2003008540-A2.
 XX PD 30-JAN-2003.
 XX PF 21-JUN-2002; 2002WO-US019668.
 XX PR 22-JUN-2001; 2001US-0300112P.
 XX PR 24-AUG-2001; 2001US-0314662P.
 XX PR 26-SEP-2001; 2001US-0325277P.
 XX PR 21-NOV-2001; 2001US-0332132P.
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
 XX PI Moughamer T, Provart N, Rickel D, Zhu T;
 XX WPI; 2003-248011/24.
 XX DT New stress-responsive nucleic acid, useful for altering the
 XX PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
 XX PT stress, salt stress or osmotic stress.
 XX PS Claim 1; SEQ ID NO 9096; 89pp; English.
 XX CC The invention relates to novel abiotic stress responsive polynucleotides
 CC and polypeptides. Also disclosed are vectors, expression cassettes, host
 CC cells, and plants containing such polynucleotides. Also disclosed are
 CC methods for using the polynucleotides and polypeptides to alter the
 CC responsiveness of a plant to abiotic stress. The invention is useful in
 CC agriculture. The nucleic acid is useful for determining whether a test
 CC plant has been exposed to an abiotic stress condition. It is also useful
 CC for selecting an agent that alters abiotic stress regulated
 CC polynucleotide expression in a plant cell, and to identify a homolog or
 CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
 CC molecule and the polypeptide encoded by it are useful in altering the
 CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
 CC stress, osmotic stress or any of their combinations. The present sequence
 CC is used in the exemplification of the invention
 XX SQ Sequence 348 AA;
 Query Match 19.6%; Score 64.5; DB 7; Length 348;
 Best Local Similarity 33.9%; Pred. No. 12;
 Matches 20; Conservative 11; Mismatches 23; Indels 5; Gaps 3;
 QY 12 IGLRTA--KINFYSLNET--GLDGVPAFKEG-VDLGAGGAIDGFTYIKNEAIIEL 65
 DB 231 VGLRVRPKFVSFYSENEVLKMGIGAFDVGKVDLSGMIXGELVWVEKVMHRAVVEV 289
 RESULT 9
 ABU33885
 ID ABU33885 standard; protein; 601 AA.
 XX AC ABU33885;
 DT 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #19412.
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Mycobacterium avium.
 XX

PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX DR N-PSDB; ACA37755.
 XX DT New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX PT isolate candidate molecules for rational drug discovery programs.
 XX PS Claim 25; SEQ ID NO 61809; 1766pp; English.
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 601 AA;
 Query Match 19.6%; Score 64.5; DB 6; Length 601;
 Best Local Similarity 32.7%; Pred. No. 23;
 Matches 17; Conservative 8; Mismatches 24; Indels 3; Gaps 1;
 QY 6 LVDDTVIGKLTAKIN---FVSLNETGLDGVPAFKEGVDLGGAGDAIDGFTY 54
 DB 257 LVADVPVCGLLSGVDSLVGLLAARQGHLSITFSGFESAGGVKGDFQY 308
 RESULT 10
 ADY05830
 ID ADY05830 standard; protein; 138 AA.
 XX

CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in
 CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.
 XX
 SQ Sequence 193 AA;
 Query Match 19.5%; Score 64; DB 8; Length 193;
 Best Local Similarity 35.0%; Pred. No. 6.6;
 Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;
 QY 6 LVDDTVIGKLTAKINFYSLNETGLDG--VPAFKEGVDLG-----GAIDQFTVHYIKN 59
 Db 102 VVDSSTDRLVTAKEEFHAILEDELKGAVLVYANKQDLPGLDAAITSELELHKIKS 161
 RESULT 14
 ADY22628
 ID ADY22628 standard; protein; 193 AA.
 XX
 AC ADY22628;
 DT 21-APR-2005 (first entry)
 XX
 DE Plant full length insert polypeptide seqid 70412.
 XX
 KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content.
 XX
 OS Unidentified.
 XX
 PN US2004034888-A1.
 XX
 PD 19-FEB-2004.
 XX
 PF 28-APR-2003; 2003US-00425114.
 XX
 PR 06-MAY-1999; 99US-00304517.
 XX
 PR 05-NOV-2001; 2001US-00985678.
 XX
 PA (LIUJ/) LIU J.
 PA (ZHOU/) ZHOU Y.
 PA (KOVA/) KOVALIC D K.
 PA (SCRE/) SCREEN S E.
 PA (TABA/) TABASKA J E.
 PA (CAOY/) CAO Y.
 XX
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 XX WPI; 2004-180133/17.
 XX
 PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.
 XX
 PS Claim 1; SEQ ID NO 70412; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp.segdata.uspto.gov/sequence.html?docid:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in
 CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.
 XX
 SQ Sequence 193 AA;
 Query Match 19.5%; Score 64; DB 8; Length 193;
 Best Local Similarity 35.0%; Pred. No. 6.6;
 Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;
 QY 6 LVDDTVIGKLTAKINFYSLNETGLDG--VPAFKEGVDLG-----GAIDQFTVHYIKN 59
 Db 102 VVDSSTDRLVTAKEEFHAILEDELKGAVLVYANKQDLPGLDAAITSELELHKIKS 161
 RESULT 15
 ADX94536
 ID ADX94536 standard; protein; 193 AA.
 XX
 AC ADX94536;
 DT 21-APR-2005 (first entry)
 XX
 DE Plant full length insert polypeptide seqid 57200.
 XX
 KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content.
 XX
 OS Unidentified.
 XX
 PN US2004034888-A1.
 XX
 PD 19-FEB-2004.
 XX
 PF 28-APR-2003; 2003US-00425114.
 XX
 PR 06-MAY-1999; 99US-00304517.
 XX
 PR 05-NOV-2001; 2001US-00985678.
 XX
 PA (LIUJ/) LIU J.
 PA (ZHOU/) ZHOU Y.
 PA (KOVA/) KOVALIC D K.
 PA (SCRE/) SCREEN S E.
 PA (TABA/) TABASKA J E.
 PA (CAOY/) CAO Y.
 XX
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 XX WPI; 2004-180133/17.
 XX
 PT New recombinant DNA construct, useful for improving plant tolerance to

PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.

XX Claim 1; SEQ ID NO 57200; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.

XX SQ Sequence 193 AA;

Query Match 19.5%; Score 64; DB 8; Length 193;

Best Local Similarity 35.0%; Pred. No. 6.6;

Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;

OY 6 LVDDTVIGKRTAKINFYSLNLETGLD--VPAPKEGVVDLAG---GAIDQFTYHYIKN 59
Db 102 VVDSSDTDLVTAKKEFHAELEDELKGAVLVYANKQDLPGALDDAAITESLELHKIKS 161

Search completed: January 24, 2006, 19:44:51
Job time : 16.5258 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 2.46592 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-18
Perfect score: 329
Sequence: 1 LAATPLVDDTVIGKLTAKI.....GAIDQFTYHYIKNEAIIEL 65
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	207	62.9	356	F70212	hypothetical prote
2	73	22.2	233	C86849	regulatory protein
3	67	20.4	337	B84335	hypothetical prote
4	64.5	19.6	307	AE0962	D-serine deaminase
5	64	19.5	284	F90016	conserved hypothet
6	63.5	19.3	301	F75419	succinyl-CoA synth
7	63.5	19.3	345	C84489	probable replicati
8	63.5	19.3	709	E90210	hypothetical prote
9	63	19.1	491	S42678	killer toxin - pic
10	62	18.8	703	D81365	carbon starvation
11	62	18.8	781	T02272	hypothetical prote
12	61	18.5	185	E65884	hypothetical prote
13	60.5	18.4	612	S61116	probable membrane
14	60	18.2	293	A64907	probable transcrip
15	60	18.2	301	S52142	glycosyl transferas
16	60	18.2	375	C87394	hypothetical prote
17	60	18.2	397	B69378	probable acyl-CoA
18	60	18.2	540	B83850	D-3-phosphoglycera
19	59	17.9	348	AD2247	hypothetical prote
20	59	17.9	402	D69530	probable acyl-CoA
21	59	17.9	1228	C98219	proline dehydrogen
22	59	17.9	1228	AG3067	proline dehydrogen
23	59	17.9	1578	S77257	hypothetical prote
24	58.5	17.8	218	S30287	regulatory protein
25	58.5	17.8	218	S07359	regulatory protein
26	58.5	17.8	309	AH3122	IS3 family transpo
27	58.5	17.8	309	G98164	IS3 family orfB pr
28	58.5	17.8	311	A65010	D-serine deaminase
29	58	17.6	182	S41938	ADP-ribosylation f

30	58	17.6	322	2	G97317	probable dehydroge
31	58	17.6	323	2	AH1470	oxidoreductase hom
32	58	17.6	589	2	B83214	probable glutamine
33	57.5	17.5	325	2	F82068	2-hydroxyacid dehy
34	57	17.3	291	2	G83660	hypothetical prote
35	57	17.3	371	2	T08759	conserved hypothet
36	57	17.3	566	2	T19653	hypothetical prote
37	57	17.3	566	2	E89722	protein G33Al1.1 [
38	57	17.3	632	2	I51682	epithelial sodium
39	57	17.3	1206	2	A64207	hypothetical prote
40	57	17.3	1274	2	S55050	cardiac myosin-bin
41	56.5	17.2	166	2	T12822	hypothetical prote
42	56.5	17.2	271	2	F95945	probable formimido
43	56.5	17.2	362	2	A37050	porin Omp2b precur
44	56.5	17.2	375	2	AC3415	porin [imported] -
45	56.5	17.2	455	2	F72326	hemolysin-related

ALIGNMENTS

RESULT 1
F70212
hypothetical protein BBA46 - Lyme disease spirochete plasmid A/lp54
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: F70212
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: F70212
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-356 <KLE>
A:Cross-references: UNIPROT:O50939; UNIPARC:UPI0000056927; GB:AE000790; NID:g2690224; PI
A:Experimental source: strain B31
C:Genetics:
C:Genome: plasmid
C:Superfamily: Lyme disease spirochete plasmid A/lp54 hypothetical protein BBA46

Query Match	62.9%	Score 207;	DB 2;	Length 356;
Best Local Similarity	66.2%	Pred. No. 3.1e-16;		
Matches	43;	Conservative	5;	Mismatches 17;
		Indels	0;	Gaps 0;
QY	1	LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAKFGVYDLGAGGAIDQFTYHYIKNE	60	
DB	198	LNAVYVYDDSLIDSIRKTNINFYSLNETGNDGILAFKEGVSLSGDPIDETFTLFYIKNE	257	
QY	61	AIIEEL 65		
DB	258	AIKEL 262		

RESULT 2

C86849
regulatory protein [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C:Accession: C86849
R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissensbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86849
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-233 <STO>
A:Cross-references: UNIPROT:O9CEP1; UNIPARC:UPI00000C6B13; GB:AE005176; PID:gl2724820; P
A:Experimental source: strain IL1403

probable replication protein A1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84489
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84489
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <STO>
A:Cross-references: UNIPROT:Q9SK24; UNIPARC:UPI000009F829; GB:AE002093; NID:G4558677; P1
C:Genetics:
A:Gene: At2g09960
A:Map position: 2

Query Match 19.3%; Score 63.5; DB 2; Length 345;
Best Local Similarity 26.4%; Pred. No. 12;
Matches 19; Conservative 10; Mismatches 20; Indels 23; Gaps 2;

Qy 2 AATPLVDDTVIGKRTAKINFY-----SLLNETGLDGVPAKGVLDL 43
Db 103 ACQQLTGIVVCLIRFAKINLYNDRSVNSFQVPSVDSSTLDELGL-----FKQSIP 157
Qy 44 AGGAIDEOFTYH 55
Db 158 DGLTLESSGSF 169

RESULT 8
E90210
hypothetical protein purL [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: E90210
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: E90210
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-709 <KUR>
A:Cross-references: UNIPROT:Q9UX24; UNIPARC:UPI0000132B43; GB:AE006641; NID:G13813798; E
C:Genetics:
A:Gene: purL
C:Superfamily: phosphoribosylformylglycinamide synthase component II

Query Match 19.3%; Score 63.5; DB 2; Length 709;
Best Local Similarity 46.5%; Pred. No. 26;
Matches 20; Conservative 5; Mismatches 17; Indels 1; Gaps 1;

Qy 5 PLVDDTVIGKRTAKINFYSLNETGLDGVPAKGVLDL 47
Db 168 PLVDDVAAGIVAKDKIK-PSIVDKAGLKLVLVLAGLTVGVDLGG 209

RESULT 9
S42678
killer toxin - Pichia acaciae plasmid pPacl-2
C:Species: Pichia acaciae
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S42678
R:Boiten, P.L.; Eastman, E.M.; Cihak, P.L.; Hayman, G.T.
Yeast 10, 403-414, 1994
A:Title: Isolation and sequence analysis of a gene from the linear DNA plasmid pPacl-2 c
A:Reference number: S42678; MUID:94287718; PMID:8017110
A:Accession: S42678

A:Molecule type: DNA
A:Residues: 1-491 <BOL>
A:Cross-references: UNIPROT:Q01498; UNIPARC:UPI000006B991; GB:U02596; NID:G435925; PIDN:
A:Experimental source: NRRL Y-18665
C:Genetics:
A:Genome: plasmid pPacl-2
C:Keywords: glycoprotein
F:1135,205,241,263,291,354,387/Binding site: carbohydrate (Asn) (covalent) #status predic
Query Match 19.1%; Score 63; DB 2; Length 491;
Best Local Similarity 25.4%; Pred. No. 20;
Matches 16; Conservative 11; Mismatches 16; Indels 20; Gaps 2;

Qy 9 DTVIG-----KLRTAKINFYSLNETGLDGVPAKGVLDL 62
Db 278 DNIYGYMQLQRIINNTNLSLTNEVTKNGVP-----DHFYFYAKNDPM 323
Qy 63 IEL 65
Db 324 IQI 326

RESULT 10
D81365
carbon starvation protein A homolog Cj0917c [similarity] - Campylobacter jejuni (strain
C:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: D81365
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: D81365
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-703 <PAR>
A:Cross-references: UNIPROT:Q9PPI6; UNIPARC:UPI000000C1D90; GB:AL139076; GB:AL111168; NID
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: cstA; Cj0917c
C:Superfamily: carbon starvation protein

Query Match 18.8%; Score 62; DB 2; Length 703;
Best Local Similarity 38.2%; Pred. No. 38;
Matches 21; Conservative 6; Mismatches 20; Indels 8; Gaps 3;

Qy 17 TAKINFYSLNETGLDGVPAKGV-----DLAGAIDEOFTYHY-INKNEAIIEL 65
Db 428 TTNIGEXTILSRGTG--GAPTFAIGVALILHELFGVLDLMAFWYHFAILFEALFIL 480

RESULT 11
T02272
hypothetical protein T1308.9 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 09-Jul-2004
C:Accession: T02272
R:Vyotskakaia, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.; Li,
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.
submitted to the EMBL Data Library, June 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.
A:Reference number: 214649
A:Accession: T02272
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-781 <VVS>
A:Cross-references: UNIPROT:O80743; UNIPARC:UPI000000A9B72; EMBL:AC004473; NID:G3108025;
C:Genetics:
A:Gene: ATSP:T13D8.9
A:Map position: 1
A:Introns: 75/3; 126/3; 259/1; 478/1; 673/3

Query Match 18.8%; Score 62; DB 2; Length 781;
Best Local Similarity 48.1%; Pred. No. 43;
Matches 13; Conservative 4; Mismatches 8; Indels 2; Gaps 1;

QY 27 NETGLDGVPAFKEGVLDLGGAI--DEQ 51
DB 177 NETGKGESKIKENIDIANSAVLTDDE 203

RESULT 12
E69884
Hypothetical protein ymcC - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: E69884
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seroz
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69884
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-185 <KUN>
A:Cross-references: UNIPROT:Q31780; UNIPARC:UPI000006042F; GB:Z99112; GB:AL009126; NID:9
A:Experimental source: strain 168
C:Genetics:
A:Gene: ymcC
C:Superfamily: Bacillus subtilis hypothetical protein ymcC

Query Match 18.5%; Score 61; DB 2; Length 185;
Best Local Similarity 28.6%; Pred. No. 11;
Matches 18; Conservative 11; Mismatches 26; Indels 8; Gaps 2;

QY 1 LAATPLVDVTGKLTAKINFYSLNLTGLDGVPAFKEGVLDLGGAI-----IDRQFTYH 55
DB 39 LALTFVID---LILLAATGVLDYRGASATAAHGIAAVVIGISYAGKQIQWADEKFOY 95

QY 56 YIK 58
DB 96 VTK 98

RESULT 13
S61116
probable membrane protein YPL070w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein LPF14w
C:Species: Saccharomyces cerevisiae
C:Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S61116
R:Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A:Reference number: S59677
A:Accession: S61116
A:Molecule type: DNA
A:Residues: 1-612 <HAL>
A:Cross-references: UNIPROT:Q02866; UNIPARC:UPI000006A497; EMBL:U41849; NID:g1147608; PI
C:Genetics:
A:Gene: MIPS:YPL070w
A:Cross-references: SGD:S0005991
A:Map position: 16L

C:Superfamily: Saccharomyces cerevisiae probable membrane protein YPL070w
C:Keywords: transmembrane protein
F:356-372/Domain: transmembrane #status predicted <TMM>

Query Match 18.4%; Score 60.5; DB 2; Length 612;
Best Local Similarity 35.0%; Pred. No. 49;
Matches 21; Conservative 10; Mismatches 18; Indels 11; Gaps 4;

QY 12 IGKLRTA-KINFYSLN-ETGLDGVPAFKEGVLDLGGAIIDRQFTYHYI---KNEAIIEL 65
DB 286 LSKLSTWKNILKFLSLNNGIDPEKPEIKDIL-----YEFTYHSISPCKEIKALKKL 340

RESULT 14
A64907
probable transcription regulator yneJ - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 05-Oct-2004
C:Accession: A64907; S41477; S35950
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64907
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-293 <BLAT>
A:Cross-references: UNIPROT:P77309; UNIPARC:UPI000013BB1C; GB:AE000250; GB:U00096; NID:9
A:Experimental source: strain K-12, substrain MGI655
R:Cecchini, E.; Dominy, P.J.; Geri, C.; Kaiser, K.; Sentry, J.; Milner, J.J.
Nucleic Acids Res. 21, 5742-5747, 1993
A:Title: Identification of genes up-regulated in dedifferentiating Nicotiana glauca pit
A:Reference number: S41477; MUID:94111998; PMID:8284223
A:Accession: S41477
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 'MP', 173-186; 190-274 <CEC>
A:Cross-references: UNIPARC:UPI0000178EDE; UNIPARC:UPI0000178EDF; EMBL:X74318; NID:94176
A:Note: the authors state that the gene has been cloned from Nicotiana glauca, most like
C:Genetics:
A:Gene: yneJ
C:Superfamily: HTH-type transcriptional regulator, LysR family, MetR type
C:Keywords: DNA binding; transcription regulation

Query Match 18.2%; Score 60; DB 2; Length 293;
Best Local Similarity 35.3%; Pred. No. 25;
Matches 12; Conservative 9; Mismatches 11; Indels 2; Gaps 1;

QY 6 LVDDTVIGKLTAKINFYSLNLTGLDGVPAFKE 39
DB 130 MLDGVLEKLNAAFID--GPINHTAIDGIPVRE 161

RESULT 15
S52142
glycosyl transferase amsB - Erwinia amylovora
C:Species: Erwinia amylovora
C:Date: 15-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004
C:Accession: S61895; S52142
R:Bugert, P.; Geiger, K.
Mol. Microbiol. 15, 917-933, 1995
A:Title: Molecular analysis of the ams operon required for exopolysaccharide synthesis o
A:Reference number: S61891; MUID:95319333; PMID:7596293
A:Accession: S61895
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <BUZ>
A:Cross-references: UNIPROT:Q46632; UNIPARC:UPI0000125A31; EMBL:X77921; NID:g600426; PID
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 18.2%; Score 60; DB 2; Length 301;
Best Local Similarity 29.3%; Pred. No. 25;

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 13.9671 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-18

Perfect score: 329

Sequence: 1 LAATPLVDDTVIGKLTAKI.....GAIDEQTYHYKNEAIIEL 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329	100.0	370	Q5XYW3	BORR
2	329	100.0	370	Q9R2W7	BORR
3	329	100.0	370	Q9S0C8	BORR
4	329	100.0	370	Q9S0C8	BORR
5	329	100.0	370	Q9S0F9	BORR
6	316	96.0	362	Q9S051	BORR
7	304	92.4	370	Q9S003	BORR
8	298	90.6	373	Q5XZ68	BORR
9	207	62.9	356	Q50939	BORR
10	126	38.3	156	Q6AS11	BORR
11	81	24.6	250	Q54MG3	DICDI
12	73.5	22.3	402	Q97BQ1	THEVO
13	73	22.2	233	MECA	LACIA
14	69.5	21.1	1046	Q8EBM6	SHEON
15	68.5	20.8	6858	Q7QUM1	ITALA
16	67	20.4	337	Q9HP24	HALSA
17	67	20.4	394	Q6LIQ8	PICTO
18	67	20.4	745	Q4X0S2	ASFPF
19	66	20.1	1592	Q5AZJ8	EMENI
20	65.5	19.9	174	Q59PV8	CANAL
21	65	19.8	284	Q5HE01	STAAC
22	65	19.8	410	Q8DV08	STRMU
23	65	19.8	445	Q90681	ASTPE
24	65	19.8	703	Q4HDM0	CAMCO
25	65	19.8	1315	Q9R9T7	PSEPU
26	65	19.8	1317	Q8BD80	PSEPK
27	64.5	19.6	307	Q5PKR5	SALPA
28	64.5	19.6	307	Q822L2	SALTI
29	64.5	19.6	307	Q8ZL10	SALTY
30	64.5	19.6	312	Q57I34	SALCH
31	64.5	19.6	601	Q740D0	MYCPA

32 64.5 19.6 1517 2 Q4HLT3 CAMLA
33 64 19.5 182 2 Q6YW27 ORISA
34 64 19.5 284 2 Q6GEM6 STAAR
35 64 19.5 284 2 Q7A485 STAAAN
36 64 19.5 284 2 Q9S666 STAAAM
37 64 19.5 589 2 Q4T1A4 AZOVI
38 63.5 19.3 248 2 Q5JXW0 HUMAN
39 63.5 19.3 301 2 Q9RUY2 DEIRA
40 63.5 19.3 345 2 Q9SK24 ARATH
41 63.5 19.3 709 1 PURL SULSO
42 63 19.1 138 2 Q73M04 TREDE
43 63 19.1 377 2 Q835P7 ENTFA
44 63 19.1 424 2 Q53KS8 ORISA
45 63 19.1 451 2 Q7XMI6 ORISA

Q4HLT3 campylobact
Q6YW27 oryza sativ
Q6GEM6 staphylococ
Q7A485 staphylococ
Q9S666 staphylococ
Q4T1A4 azotobacter
Q5JXW0 homo sapien
Q9RUY2 deinococcus
Q9SK24 arabidopsis
PURL sulfolobus
Q73M04 treponema d
Q835P7 enterococu
Q53KS8 oryza sativ
Q7XMI6 oryza sativ

ALIGNMENTS

RESULT 1
Q5XYW3 BORGA PRELIMINARY; PRT; 370 AA.
AC Q5XYW3;
DT 25-OCT-2004 (Tremblrel. 28, Created)
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames-BGPi68;
OS Borrelia garinii PBI;
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Glockner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Schilhabel M., Wilke B., Suhnel J., Platzner M.;
RT "Comparative analysis of the Borrelia garinii genome."
RL Nucleic Acids Res. 32:6038-6046(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Schilhabel M., Lehmann R., Platzner M.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY722923; AAU86019.1; -; Genomic_DNA.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
KW Hypothetical protein.
SQ SEQUENCE 370 AA; 42173 MW; 62FB9015CEBFC9C3 CRC64;

Query Match 100.0%; Score 329; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 6.6e-30;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKGVLDLAGAIDQFTYHYKNE 60
Db 211 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKGVLDLAGAIDQFTYHYKNE 270

QY 61 AIIEL 65
Db 271 AIIEL 275

RESULT 2

Q9R2W7 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9R2W7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-FEB-2005 (Tremblrel. 29, Last annotation update)
DE Hypothetical protein BB11 (Hypothetical protein BBP11) (Hypothetical protein BBS11).
GN OrderedLocusNames=BB11, BBP11, BBS11;

OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-8, plasmid cp32-1, and plasmid cp32-3.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=cp32-8, cp32-1, and cp32-3;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RA "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=cp32-8, cp32-1, and cp32-3;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RA "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AF001575; AAF07434.1; -; Genomic DNA.
DR EMBL; AF001576; AAF07479.1; -; Genomic DNA.
DR EMBL; AF001580; AAF07651.1; -; Genomic DNA.
DR TIGR; BB11; -.
DR TIGR; BBP11; -.
DR TIGR; BBS11; -.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
DR Complete proteome; Hypothetical protein; Plasmid.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 370 AA; 42102 MW; D071E1723FA9767C CRC64;

Query Match 100.0%; Score 329; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 6.6e-30;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAPKGVDLGAGADEQFTYHYKNE 60
Db 211 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAPKGVDLGAGADEQFTYHYKNE 270

QY 61 AIIEL 65
Db 271 AIIEL 275

RESULT 3
Q9S096 BORBU
ID Q9S096 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9S096;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BB011.
GN OrderedLocustNames=BB011.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-7.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RA "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,

RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RA "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RA "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AF001575; AAF07607.1; -; Genomic DNA.
DR TIGR; BBO11; -.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
DR Complete proteome; Hypothetical protein; Plasmid.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 370 AA; 42103 MW; 62FB9015CFEBBCD6 CRC64;

Query Match 100.0%; Score 329; DB 2; Length 370;
Best Local Similarity 100.0%; Pred. No. 6.6e-30;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAPKGVDLGAGADEQFTYHYKNE 60
Db 211 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAPKGVDLGAGADEQFTYHYKNE 270

QY 61 AIIEL 65
Db 271 AIIEL 275

RESULT 4
Q9S0C8 BORBU
ID Q9S0C8 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9S0C8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BBM11.
GN OrderedLocustNames=BBM11.
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-6.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RA "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,

RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
 RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
 RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
 RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
 RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
 RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.,
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
 RL Nature 390:580-586(1997).
 DR EMBL; AE001578; AAF07566.1; -; Genomic_DNA.
 DR TIGR; BBN11; -.
 DR InterPro; IPR008505; DUF787.
 DR Pfam; PF05619; DUF787; 1.
 KW Complete proteome; Hypothetical protein; Plasmid.
 SQ SEQUENCE 370 AA; 42158 MW; 62FEE72EB97CF9C3 CRC64;

 Query Match 100.0%; Score 329; DB 2; Length 370;
 Best Local Similarity 100.0%; Pred. No. 6.6e-30;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGGAIDQFTYHYIKNE 60
 |||||
 DB 211 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGGAIDQFTYHYIKNE 270

 QY 61 AIIEL 65
 |||||
 DB 271 AIIEL 275

 RESULT 5
 Q9S0F9 BORBU
 ID Q9S0F9 BORBU PRELIMINARY; PRT; 370 AA.
 AC Q9S0F9
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE Hypothetical protein BBR11.
 GN OrderedLocusNames=BBR11;
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OG Plasmid cp32-4.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 OC Borrelia burgdorferi group.
 OX NCBI_TaxID=139;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=10672174;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
 RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
 RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
 RA "A bacterial genome in flux: the twelve linear and nine circular
 RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
 RT spirochete Borrelia burgdorferi."
 RL Mol. Microbiol. 35:490-516(2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=10672174;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
 RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
 RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
 RA "A bacterial genome in flux: the twelve linear and nine circular
 RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
 RT spirochete Borrelia burgdorferi."
 RL Mol. Microbiol. 35:490-516(2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=10672174;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
 RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
 RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
 RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
 RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
 RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
 RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
 RL Nature 390:580-586(1997).
 DR EMBL; AE001577; AAF07520.1; -; Genomic_DNA.
 DR TIGR; BBR11; -.
 DR InterPro; IPR008505; DUF787.
 DR Pfam; PF05619; DUF787; 1.

KW Complete proteome; Hypothetical protein; Plasmid.
 SQ SEQUENCE 370 AA; 42130 MW; BE37D9DB14A1B0BD CRC64;

 Query Match 100.0%; Score 329; DB 2; Length 370;
 Best Local Similarity 100.0%; Pred. No. 6.6e-30;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGGAIDQFTYHYIKNE 60
 |||||
 DB 211 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGGAIDQFTYHYIKNE 270

 QY 61 AIIEL 65
 |||||
 DB 271 AIIEL 275

 RESULT 6
 Q9S051 BORBU
 ID Q9S051 BORBU PRELIMINARY; PRT; 362 AA.
 AC Q9S051
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE Hypothetical protein BEN11.
 GN OrderedLocusNames=BBN11;
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OG Plasmid cp32-9.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
 OC Borrelia burgdorferi group.
 OX NCBI_TaxID=139;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=20138354; PubMed=10672174;
 RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
 RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
 RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
 RA "A bacterial genome in flux: the twelve linear and nine circular
 RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
 RT spirochete Borrelia burgdorferi."
 RL Mol. Microbiol. 35:490-516(2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 35210 / B31;
 RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
 RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
 RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
 RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
 RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
 RA Artach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
 RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi."
 RL Nature 390:580-586(1997).
 DR EMBL; AE001581; AAF07685.1; -; Genomic_DNA.
 DR TIGR; BBN11; -.
 DR InterPro; IPR008505; DUF787.
 DR Pfam; PF05619; DUF787; 1.
 KW Complete proteome; Hypothetical protein; Plasmid.
 SQ SEQUENCE 362 AA; 41367 MW; CD76E4F607088587 CRC64;

 Query Match 96.0%; Score 316; DB 2; Length 362;
 Best Local Similarity 95.4%; Pred. No. 2.1e-28;
 Matches 62; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

 QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGGAIDQFTYHYIKNE 60
 |||||
 DB 211 LAATPLVDDTVIGKLTAKINFYSLNETGLGVPAFKEGVDLAGSTIDEQFTYHYIKNE 270

 QY 61 AIIEL 65
 |||||

Db 271 AIIEL 275

RESULT 7

Q9S003 BORBU

ID Q9S003 BORBU PRELIMINARY; PRT; 370 AA.

AC Q9S003; 271 AIIEL 275

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein BBQ18.

GN OrderedLocNames=BBQ18;

OS Borrelia burgdorferi (Lyme disease spirochete).

OG Plasmid lp56.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;

OC Borrelia burgdorferi group.

OX NCBI_TaxID=139;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ATCC 35210 / B31;

RC MEDLINE=20138354; PubMed=10672174;

RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,

RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,

RA Hickey E.K., Gwinn M.L., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,

RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,

RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,

RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,

RA Gocayne J.D., Weidman J.F., Utterback T.R., Wattney L., McDonald L.A.,

RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,

RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.,

RT "Genomic sequence of a Lyme disease spirochete, Borrelia

RT burgdorferi."

RL Nature 390:580-586 (1997).

RL EMBL; AR001584; AAF07741.1; -; Genomic_DNA.

DR TIGR; BBQ18; -.

DR InterPro: IPR008505; DUF787.

DR Pfam; PF05619; DUF787; 1.

KW Complete proteome; Hypothetical protein; Plasmid.

SQ SEQUENCE 370 AA; 42150 MW; DAEE039E8F8EBD8 CRC64;

Query Match 92.4%; Score 304; DB 2; Length 370;

Best Local Similarity 92.3%; Pred. No. 5.6e-27;

Matches 60; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LAATPLVDVTGKLTAKINFYSLNLTGDPAPKFGVDLAGAIDEQFTYHYKNE 60

Db 211 LAATPLVDVTGKLTAKINFYSLNLTGDPAPKFGVDLAGAIDEQFTYHYKNE 270

QY 61 AIIEL 65

Db 271 AIIEL 275

RESULT 8

Q9X268 BORG

ID Q9X268 BORG PRELIMINARY; PRT; 373 AA.

AC Q9X268; 271 AIIEL 275

DT 25-OCT-2004 (TrEMBLrel. 28, Created)

DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Hypothetical protein.

GN ORFNames=BGFP064;

OS Borrelia garinii PBI.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;

OC Borrelia burgdorferi group.

OX NCBI_TaxID=290434;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=PBI;

RA Glockner G., Lehmann R., Romualdi A., Pradella S.,

RA Schloker-Spechtel U., Schilhabel M., Wilske B., Suhnel J., Platzer M.;

RT "Comparative analysis of the Borrelia garinii genome."

RL Nucleic Acids Res. 32:6038-6046(2004).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=PBI;

RA Gloeckner G., Schilhabel M., Lehmann R., Platzer M.;

RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY722919; AAU85914.1; -; Genomic_DNA.

DR InterPro: IPR008505; DUF787.

DR Pfam; PF05619; DUF787; 1.

KW Hypothetical protein.

SQ SEQUENCE 373 AA; 42512 MW; FC40FB96B75ED69 CRC64;

Query Match 90.6%; Score 298; DB 2; Length 373;

Best Local Similarity 89.2%; Pred. No. 2.8e-26;

Matches 58; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LAATPLVDVTGKLTAKINFYSLNLTGDPAPKFGVDLAGAIDEQFTYHYKNE 60

Db 214 LTAAPLVDVTGKLTAKINFYSLNLTGDPAPKFGVDLAGAIDEQFTYHYKNE 273

QY 61 AIIEL 65

Db 274 AIIEL 278

RESULT 9

OS0939 BORBU

ID OS0939 BORBU PRELIMINARY; PRT; 356 AA.

AC OS0939;

DT 01-JUN-1998 (TrEMBLrel. 06, Created)

DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein BBA46.

GN OrderedLocNames=BBA46;

OS Borrelia burgdorferi (Lyme disease spirochete).

OG Plasmid lp54.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;

OC Borrelia burgdorferi group.

OX NCBI_TaxID=139;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ATCC 35210 / B31;

RC MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,

RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,

RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,

RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,

RA Gocayne J.D., Weidman J.F., Utterback T.R., Wattney L., McDonald L.A.,

RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,

RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.,

RT "Genomic sequence of a Lyme disease spirochete, Borrelia

RT burgdorferi."

RL Nature 390:580-586 (1997).

RL EMBL; AE000790; AAC66278.1; -; Genomic_DNA.

DR PIR; F70212; F70212.

DR TIGR; BBA46; -.

DR InterPro: IPR008505; DUF787.

DR Pfam; PF05619; DUF787; 1.

KW Complete proteome; Hypothetical protein; Plasmid.

SQ SEQUENCE 356 AA; 40359 MW; A1343C1F56ED474 CRC64;

Query Match 62.9%; Score 207; DB 2; Length 356;

Best Local Similarity 66.2%; Pred. No. 1.3e-15;

```

Matches 43; Conservative 5; Mismatches 17; Indels 0; Gaps 0;
QY 1 LAATPLVDVTGKLTAKINFYSLNLTGDPVAFKGVLDAGAIHQFTYHYKNE 60
Db 198 LNAVVPYDSDLSLRKTNINFYSLNLTGNDGILAFKGVSLSGDPIDETFTLFYKNE 257
QY 61 AIIEL 65
Db 258 AIKEL 262

RESULT 10
QASII_BORGA
ID Q6ASII_BORGA PRELIMINARY; PRT; 156 AA.
AC Q6ASII;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=BGA44;
OS Borrelia garinii.
OG Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Stuehn J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000015; AAT93803.1; -; Genomic_DNA.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 156 AA; 17537 MW; 97310D5A88A9CF6B CRC64;

Query Match 38.3%; Score 126; DB 2; Length 156;
Best Local Similarity 65.8%; Pred. No. 1.6e-06;
Matches 25; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
QY 1 LAATPLVDVTGKLTAKINFYSLNLTGDPVAFK 38
Db 119 LNSTPYDSDLSLRKTNINFYSLNLTGNDGILAFK 156

RESULT 11
Q54MG9_DICDI
ID Q54MG9_DICDI PRELIMINARY; PRT; 250 AA.
AC Q54MG9;
DT 13-SEP-2005 (TReMBLrel. 31, Created)
DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=DD0218760;
OS Dictyostellum discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero P.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsey R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,

```

```

RA Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaalsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AAF10100127; EAL64517.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 250 AA; 29183 MW; 16D29CAF2BE5080F CRC64;

Query Match 24.6%; Score 81; DB 2; Length 250;
Best Local Similarity 48.6%; Pred. No. 0.5;
Matches 18; Conservative 5; Mismatches 12; Indels 2; Gaps 1;
QY 31 LDGVPAPKGVLDAGAI--DEQFTYHYKNEAIIEL 65
Db 150 LDGLQVFKGVCLYCGSIDCEDFTFYCYCNSTLSI 186

RESULT 12
Q97BQ1_THEVO
ID Q97BQ1_THEVO PRELIMINARY; PRT; 402 AA.
AC Q97BQ1;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Aspartate aminotransferase.
GN OrderedLocNames=TW0404; ORFNames=TVG0393535;
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasma; Thermoplasmales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031; DOI=10.1073/pnas.97.26.14257;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunohiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; BA000011; BAB59546.1; -; Genomic_DNA.
DR GO; GO:0008483; P:transaminase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR004839; AminoTrans I/II.
DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF0155; AminoTran_1_2; I.
DR PROSITE; PS01048; RIBOSOMAL_S6; UNKNOWN_1.
KW AminoTransferase; Complete proteome; Transferase.
SQ SEQUENCE 402 AA; 45684 MW; 1CE50D5AB95AB5DA CRC64;

Query Match 22.3%; Score 73.5; DB 2; Length 402;
Best Local Similarity 24.1%; Pred. No. 6.3;
Matches 21; Conservative 12; Mismatches 25; Indels 29; Gaps 2;
QY 5 PLVDDTVTGKLR-----TAKINFYSLNLTG-----DGVP 35
Db 204 PIVENPYQLRYDGEFVQSIKMDKNGNVIYLGTFKWMAPGLRGLYVIAPREVIDKIN 263
QY 36 AFKEGVDLAGGAIHQFTYHYKNEAI 62
Db 264 LLKQGLDLSDSLSEYIAVEYKRGAI 290

RESULT 13
MECA_LACIA

```

RA	Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA	Utterback T.R., McDonald L.A., Feidblyum T.V., Smith H.O.,
RA	Venter J.C., Nealson K.H., Fraser C.M.;
RT	"Genome sequence of the dissimilatory metal ion-reducing bacterium
RT	Shewanella oneidensis.";
RLL	Nat. Biotechnol. 20:1118-1123(2002).
DR	EMB3; AE015784; AAN56477.1; -; Genomic_DNA.
DR	HSSP; F31224; 1IWG.
DR	TIGR; SO3484; -.
DR	InterPro; IPR001036; Acrflvin_res.
DR	Pfam; PF00873; ACR tran; 1.
DR	PRINTS; PR00702; ACRIFLAVINRP.
KW	Complete proteome.
SQL	SEQUENCE 1046 AA; 113938 MW; 49BI7826F50EEED CRC64;

Query Match	21.1%;	Score 69.5;	DB 2;	Length 1046;
Best Local Similarity	31.2%;	Pred. No. 54;		
Matches 20;	Conservative 15;	Mismatches 20;	Indels 9;	Gaps 3;

QY	8	DDTVIGKLR	TAKI-----NFYSLLN	ETGLDGV	PAPF-KEGV	DLAGGAID	EQFTVHY	IKNEA 61
	:	:	:	:	:	:	:	:
	:	:	:	:	:	:	:	:
DB	221	DDTTVMSV	RTKRLYTPK	DFDYL	VVRTASD	GFPI	YLVKDVAD	VAVGAQENSTF---KSDG 277

QY	62	IIEFL 65
	:	:
DB	278	IVNL 281

RESULT 15
Q7QWU1_GIALA
ID Q7QWU1_GIALA PRELIMINARY; PRT; 6858 AA.
AC Q7QWU1;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
GLP 561 28478..7902.
DE Giardia lamblia ATCC 50803.
OS Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OC NCBI_TaxID=184922;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=WB C6;
RC Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA Olsen G.J., Sogin M.L.;
RA "Draft sequence of the Giardia lamblia genome.";
RT Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACB01000089; EAA38820.1; -; Genomic_DNA.
DR InterPro; IPR001304; Lectin C.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000264; Serum_albumin.
DR InterPro; IPR005829; Sug_transporter.
DR PROSITE; PS00212; ALBUMIN; UNKNOWN 1.
DR PROSITE; PS00615; C TYPE LECTIN 1; UNKNOWN 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQL SEQUENCE 6858 AA; 767985 MW; 0B6D765D89303D7E CRC64;

Query Match	20.8%;	Score 68.5;	DB 2;	Length 6858;
Best Local Similarity	28.8%;	Pred. No. 5.6e+02;		
Matches 19;	Conservative 8;	Mismatches 22;	Indels 17;	Gaps 1;

QY	8	DDTVIGKLR	TAKINFYSLN	ETGLDGV	PAPFKEGV	DL-----AGGAID	E 50	
	:	:	:	:	:	:	:	
	:	:	:	:	:	:	:	
DB	1754	DISSIGAIRT	PTTELFYYLLN	QTNLE	GFDPV	FEAPIG	STGSVACRHR	YEFFINALQASMTID 1813

QY	51	QFTYH 56
	:	:
DB	1814	ESLIFY 1819

Search completed: January 24, 2006, 19:56:58
Job time : 15.9671 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 4.00711 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-18
Perfect score: 329
Sequence: 1 LAAPPLVDVTGKLTAKI.....GAIDEQFTYHYKNEAIJEL 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65.5	19.9	214	2	US-09-248-796A-17440
2	63	19.1	381	2	US-09-134-000C-5229
3	63	19.1	491	2	US-09-029-267-2
4	62	18.8	351	2	US-09-902-540-12944
5	61.5	18.7	302	2	US-09-710-279-2980
6	61.5	18.7	307	2	US-09-270-767-44911
7	61.5	18.7	327	2	US-09-489-039A-10690
8	61.5	18.7	350	2	US-09-134-001C-3383
9	61	18.5	371	2	US-09-134-000C-5173
10	60.5	18.4	612	2	US-09-538-092-775
11	60.5	18.4	645	2	US-09-252-991A-23253
12	59.5	18.1	137	2	US-09-710-279-530
13	58	17.6	647	2	US-09-252-991A-17460
14	58	17.6	1326	2	US-09-489-039A-7584
15	57	17.3	568	2	US-09-628-966-10
16	57	17.3	969	2	US-09-949-016-8059
17	56.5	17.2	274	2	US-09-248-796A-17624
18	56.5	17.2	5085	2	US-09-107-532A-5087
19	56.5	17.2	529	2	US-09-248-796A-15199
20	56.5	17.2	769	2	US-09-543-681A-7175
21	56	17.0	190	2	US-09-902-540-9718
22	56	17.0	202	2	US-09-252-991A-17985
23	56	17.0	331	2	US-09-248-796A-26542
24	56	17.0	1038	2	US-09-538-092-487
25	55.5	16.9	244	2	US-09-252-991A-23692
26	55.5	16.9	496	2	US-09-252-991A-26781
27	55	16.7	392	2	US-09-583-110-3805

28	55	16.7	410	2	US-09-107-433-5135	Sequence 5135, Ap
29	55	16.7	450	2	US-09-252-991A-25760	Sequence 25760, A
30	54.5	16.6	400	2	US-09-352-990-18	Sequence 18, Appl
31	54.5	16.6	413	2	US-10-360-101-249	Sequence 249, App
32	54.5	16.6	444	1	US-08-559-260-2	Sequence 2, Appli
33	54	16.4	216	2	US-09-270-767-45064	Sequence 45064, A
34	54	16.4	362	2	US-09-252-991A-24127	Sequence 24127, A
35	54	16.4	666	2	US-09-252-991A-32542	Sequence 32542, A
36	54	16.4	2736	2	US-09-252-991A-30227	Sequence 30227, A
37	53.5	16.3	187	2	US-09-673-395A-369	Sequence 369, App
38	53.5	16.3	519	2	US-09-134-000C-4563	Sequence 4563, Ap
39	53.5	16.3	627	2	US-09-328-352-4917	Sequence 4917, Ap
40	53.5	16.3	630	2	US-09-328-352-7722	Sequence 7722, Ap
41	53.5	16.3	1318	2	US-09-540-236-3623	Sequence 3623, Ap
42	53	16.1	117	2	US-09-540-236-2827	Sequence 2827, Ap
43	53	16.1	242	2	US-09-605-703B-148	Sequence 148, App
44	53	16.1	242	2	US-09-605-703B-150	Sequence 150, App
45	53	16.1	255	2	US-09-543-681A-6904	Sequence 6904, Ap

ALIGNMENTS

RESULT 1
US-09-248-796A-17440
; Sequence 17440, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17440
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17440

Query Match 19.9%; Score 65.5; DB 2; Length 214;

Best Local Similarity 27.5%; Pred. No. 0.65;

Matches 19; Conservative 11; Mismatches 22; Indels 17; Gaps 3;

QY 12 IKGKLTAKINFVSLNETGLDGVPAF-----KEGVDLAGGAIDQFTYHY---56

Db 44 VAKQAARKVDNFKNVGLGLTGTASSLTAFKRRHDEAKKEFIDLSQSSTIDDFN-HYRS 102

QY 57 IKNEAIE 64

Db 103 ILKNSKVD 111

RESULT 2

US-09-134-000C-5229

; Sequence 5229, Application US/09134000C

; Patent No. 6617156

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1

```

; SEQ ID NO 5229
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5229

Query Match      19.1%; Score 63; DB 2; Length 381;
Best Local Similarity 34.4%; Pred. No. 3.1;
Matches 22; Conservative 10; Mismatches 22; Indels 10; Gaps 3;

QY 10 TWIGKLTAKINFYSLNETGLD-----GVPAPKEGVDLGAGAIDE-QFTYHYIKNEA 61
Db 185 TVVAGRAINEMN--SNINVTYDKKIIDVNDVVELTEGIDIIVCAIDRPPFLIHRIVNEA 242

QY 62 IIEL 65
Db 243 IVKV 246

RESULT 3
US-09-029-267-2
; Sequence 2, Application US/09029267
; Patent No. 6107057
; GENERAL INFORMATION:
; APPLICANT: Crawford, Kenneth
; APPLICANT: Zaror, Isabel
; APPLICANT: Innis, Michael
; TITLE OF INVENTION: Pichia Secretory Leader for Protein
; TITLE OF INVENTION: Expression
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: United States
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/029,267
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Chung, Ling-Fong
; REGISTRATION NUMBER: 36,482
; REFERENCE/DOCKET NUMBER: 1165.100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 601-2704
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 491 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-029-267-2

Query Match      19.1%; Score 63; DB 2; Length 491;
Best Local Similarity 25.4%; Pred. No. 4.3;
Matches 16; Conservative 11; Mismatches 16; Indels 20; Gaps 2;

QY 9 DTVIG-----KLRTAKINFYSLNETGLDGVPAKKEGVDLGAGAIDEQFTYHYIKNEAI 62
Db 278 DNIVGVMMGLQRINNTENLNFSLTNEVTKNGVP-----DHFFYYEYAKNDPM 323

QY 63 IEL 65
Db 324 IQI 326

```

```

RESULT 4
US-09-902-540-12944
; Sequence 12944, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wisegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12944
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12944

Query Match      18.8%; Score 62; DB 2; Length 351;
Best Local Similarity 37.0%; Pred. No. 3.8;
Matches 20; Conservative 6; Mismatches 14; Indels 14; Gaps 4;

QY 7 VDDTVIGKLTAKINFYSLNETGLDGVPAKKEGVDLGAGAIDEQFT---YHYI 57
Db 290 VDLTYLGHIIH---SYTSFANA---GIPAFISG---GGGAIPETFDGVRHYL 332

RESULT 5
US-09-710-279-2980
; Sequence 2980, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2980
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-2980

Query Match      18.7%; Score 61.5; DB 2; Length 302;
Best Local Similarity 32.1%; Pred. No. 3.7;
Matches 18; Conservative 10; Mismatches 17; Indels 11; Gaps 2;

QY 4 TPLVDDTVIGKLTAKINFYSLNETGLDGVPAKKEGVDLGAGAIDEQFTYHYIKN 59
Db 160 TPLAEFNIYCDEPAQAQIVFNS-----GLPLTWIGLDLAREAL---FTTHFVKD 204

RESULT 6
US-09-270-767-44911
; Sequence 44911, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094

```



```
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 44911
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44911

Query Match      18.7%; Score 61.5; DB 2; Length 307;
Best Local Similarity 30.4%; Pred. No. 3.8;
Matches 14; Conservative 9; Mismatches 20; Indels 3; Gaps 1;

QY 20 INFYSLNLTGDLGVPAPKFGVDLAGGAIDQFTYHYIKNEAII 65
Db 193 LDFFHFLNNINVTALTP---YHIAGASIDQQLQHVAKTVAMVNL 235

RESULT 7
US-09-489-039A-10690
; Sequence 10690, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10690
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10690

Query Match      18.7%; Score 61.5; DB 2; Length 327;
Best Local Similarity 37.2%; Pred. No. 4.1;
Matches 16; Conservative 8; Mismatches 16; Indels 3; Gaps 1;

QY 24 SILNLTGDLGVPAPKFGVDLA---GGAIDQFTYHYIKNEAII 63
Db 153 SILTVLTGNDVNLQAGIDLAIYFDDAPSSQLSHFLMDAIV 195

RESULT 8
US-09-134-001C-3383
; Sequence 3383, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3383
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3383

Query Match      18.7%; Score 61.5; DB 2; Length 350;
```

```
Best Local Similarity 32.1%; Pred. No. 4.5;
Matches 18; Conservative 10; Mismatches 17; Indels 11; Gaps 2;

QY 4 TPLVDDTVIGKLRITAKINFYSLNLTGDLGVPAPKFGVDLAGGAIDQFTYHYIKN 59
Db 208 TPLAEFNIYCDPEAAQIVFNS-----GLPLTWIGLDLAREAL---FTTHFVKD 252

RESULT 9
US-09-134-000C-5173
; Sequence 5173, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5173
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5173

Query Match      18.5%; Score 61; DB 2; Length 371;
Best Local Similarity 41.2%; Pred. No. 5.7;
Matches 14; Conservative 6; Mismatches 10; Indels 4; Gaps 1;

QY 34 VPAFKEGVDL----AGGAIDQFTYHYIKNEAII 63
Db 76 VPESFEGIDLALFSAGGSISEKFAPEAVKRGAVV 109

RESULT 10
US-09-538-092-775
; Sequence 775, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqformatter Version 0.9
; SEQ ID NO 775
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPL070W
US-09-538-092-775

Query Match      18.4%; Score 60.5; DB 2; Length 612;
Best Local Similarity 35.0%; Pred. No. 13;
Matches 21; Conservative 10; Mismatches 18; Indels 11; Gaps 4;

QY 12 IGLKRTA-KINFYSLN-ETGLDGVPAFKFGVDLAGGAIDQFTYHYI----KNEAII 65
Db 286 LSKUSTMKNILNLFSLNGLIDPEKFEIKDIL-----YFTYHSISPCFKIKALLKL 340
```

```

RESULT 11
US-09-252-991A-25253
; Sequence 25253, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25253
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25253
Query Match 18.4%; Score 60.5; DB 2; Length 645;
Best Local Similarity 36.2%; Pred. No. 14;
Matches 17; Conservative 8; Mismatches 17; Indels 5; Gaps 2;

QY 11 VIGKLRTA-KINFYSLNETGLDGVPAKFGVDLAGGAIDQFTYHY 56
DB 114 VTNKLEATSIHWHGILPYQMDGVP-----GISFNGIAPGETTYRF 156

RESULT 12
US-09-710-279-530
; Sequence 530, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR FILING DATE: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 530
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-530
Query Match 18.1%; Score 59.5; DB 2; Length 137;
Best Local Similarity 33.9%; Pred. No. 2.5;
Matches 21; Conservative 9; Mismatches 23; Indels 9; Gaps 3;

QY 7 VDDTVIGKLRKAKINFYSLNETGLDGVPAKFGVDLAGGAI-----DEQFTYHYIKNEAI 62
DB 16 VNDGSIYGITKDLVRDYVLMSQTLKND----KNGVEIYEGDIIEPDESFQCFPF-DDEAI 70

QY 63 IE 64
DB 71 VE 72

RESULT 13
US-09-252-991A-17460
; Sequence 17460, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

```

```

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17460
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17460
Query Match 17.6%; Score 58; DB 2; Length 647;
Best Local Similarity 40.0%; Pred. No. 31;
Matches 12; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 25 LLNETGLDGVPAKFGVDLAGGAIDQFTY 54
DB 337 LLHEAGVDNLLTFSIGFEDAGGERGDFQY 366

RESULT 14
US-09-489-039A-7584
; Sequence 7584, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 7584
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7584
Query Match 17.6%; Score 58; DB 2; Length 1326;
Best Local Similarity 33.9%; Pred. No. 79;
Matches 21; Conservative 6; Mismatches 19; Indels 16; Gaps 3;

QY 13 GKLTAKINFYSLNET-----GLDGVPAKFGVDLAGGAIDQFTYHYIKNEAI 62
DB 218 GLLFTGKL--VSTHNETSLRSLNRIGKSGEPLIRKGVDMARLNGEQF-----VTGETI 271

QY 63 IE 64
DB 272 AE 273

RESULT 15
US-09-628-966-10
; Sequence 10, Application US/09628966
; Patent No. 6596527
; GENERAL INFORMATION:
; APPLICANT: YEH, EDWARD T.H.
; APPLICANT: GONG, LIMIN
; TITLE OF INVENTION: COMPOSITION AND METHODS RELATING TO SENP1 - A
; FILE REFERENCE: UTSH-245U1
; CURRENT APPLICATION NUMBER: US/09/628,966
; CURRENT FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/146,774
; PRIOR FILING DATE: 1999-07-31

```


This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 11.4241 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-18
Perfect score: 329
Sequence: 1 LAATPLVDVTGKLTAKINFYSLNLTGLDGVPAFKGVDLGAGDAIDQFTYHYKNE 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	329	100.0	65	US-10-688-058-18	Sequence 18, Appl
2	329	100.0	160	US-10-688-058-62	Sequence 62, Appl
3	329	100.0	370	US-10-688-058-20	Sequence 20, Appl
4	329	100.0	370	US-10-688-058-64	Sequence 64, Appl
5	67	20.4	182	US-10-424-599-151311	Sequence 151311,
6	67	20.4	188	US-10-739-930-10002	Sequence 10002, A
7	65	19.8	445	US-10-732-923-2799	Sequence 2799, Ap
8	64.5	19.6	601	US-10-282-122A-61809	Sequence 61809, A
9	64	19.5	100	US-10-767-701-32620	Sequence 32620, A
10	64	19.5	124	US-10-425-115-210704	Sequence 210704,
11	64	19.5	138	US-10-425-114-61645	Sequence 61645, A
12	64	19.5	148	US-10-739-930-8484	Sequence 8484, Ap
13	64	19.5	182	US-10-437-963-187208	Sequence 187208,
14	64	19.5	193	US-10-425-114-48053	Sequence 48053, A
15	64	19.5	193	US-10-425-114-57200	Sequence 57200, A
16	64	19.5	193	US-10-425-114-60717	Sequence 60717, A
17	64	19.5	193	US-10-425-114-70412	Sequence 70412, A
18	63.5	19.3	301	US-10-369-493-576	Sequence 576, App
19	63.5	19.3	444	US-10-424-599-271469	Sequence 271469,
20	63	19.1	380	US-10-282-122A-57349	Sequence 57349, A
21	63	19.1	489	US-10-424-599-237607	Sequence 237607,
22	63	19.1	491	US-10-179-046-2	Sequence 2, Appli
23	62.5	19.0	402	US-10-369-493-18252	Sequence 18252, A
24	62.5	19.0	422	US-10-437-963-103770	Sequence 103770,
25	62.5	19.0	1017	US-10-369-493-20166	Sequence 20166, A
26	62	18.8	162	US-10-767-701-54228	Sequence 54228, A
27	62	18.8	750	US-10-171-404A-2	Sequence 2, Appli

28	61.5	18.7	122	4	US-10-282-122A-52306	Sequence 52306, A
29	61.5	18.7	302	4	US-10-282-122A-71221	Sequence 71221, A
30	61.5	18.7	350	4	US-10-724-972A-5476	Sequence 5476, Ap
31	61.5	18.7	457	4	US-10-156-761-13148	Sequence 13148, A
32	61.5	18.7	695	5	US-10-450-763-59194	Sequence 59194, A
33	61	18.5	355	3	US-09-815-242-10739	Sequence 10739, A
34	61	18.5	355	4	US-10-282-122A-57010	Sequence 57010, A
35	61	18.5	355	5	US-10-953-901-383	Sequence 383, App
36	61	18.5	355	5	US-10-953-901-385	Sequence 385, App
37	61	18.5	384	3	US-09-738-626-6407	Sequence 6407, Ap
38	61	18.5	388	4	US-10-781-014-414	Sequence 414, App
39	61	18.5	388	4	US-10-781-014-416	Sequence 416, App
40	60	18.2	144	3	US-09-925-298-534	Sequence 534, App
41	60	18.2	144	4	US-10-102-806-534	Sequence 14666, A
42	60	18.2	329	4	US-10-156-761-14666	Sequence 1013, Ap
43	60	18.2	397	4	US-10-369-493-1013	Sequence 67684, A
44	60	18.2	596	4	US-10-282-122A-67684	Sequence 13693, A
45	60	18.2	1317	4	US-10-369-493-13693	

ALIGNMENTS

RESULT 1

US-10-688-058-18
; Sequence 18, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-18

Query Match 100.0%; Score 329; DB 5; Length 65;
Best Local Similarity 100.0%; Pred. No. 4.5e-37;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LAATPLVDVTGKLTAKINFYSLNLTGLDGVPAFKGVDLGAGDAIDQFTYHYKNE	60
DB	1	LAATPLVDVTGKLTAKINFYSLNLTGLDGVPAFKGVDLGAGDAIDQFTYHYKNE	60

RESULT 2

US-10-688-058-62
; Sequence 62, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-62

Query Match          100.0%; Score 329; DB 5; Length 160;
Best Local Similarity 100.0%; Pred. No. 1.5e-36;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 60
   |||||
Db 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 60
   |||||

QY 61 AIIEL 65
   |||||
Db 271 AIIEL 275
   |||||

RESULT 3
US-10-688-058-20
; Sequence 20, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; FILE REFERENCE: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-20

Query Match          100.0%; Score 329; DB 5; Length 370;
Best Local Similarity 100.0%; Pred. No. 4.6e-36;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 60
   |||||
Db 211 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 270
   |||||

QY 61 AIIEL 65
   |||||
Db 271 AIIEL 275
   |||||

RESULT 4
US-10-688-058-64
; Sequence 64, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; FILE REFERENCE: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-64

Query Match          100.0%; Score 329; DB 5; Length 370;
Best Local Similarity 100.0%; Pred. No. 4.6e-36;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 60
   |||||
Db 211 LAATPLVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAGGAIDEQFTYHYIKNE 270
   |||||

QY 61 AIIEL 65
   |||||
Db 271 AIIEL 275
   |||||

RESULT 5
US-10-424-599-151311
; Sequence 151311, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 151311
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_107657C.1.pep
US-10-424-599-151311

Query Match          20.4%; Score 67; DB 4; Length 182;
Best Local Similarity 35.0%; Pred. No. 1.3;
Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;

QY 6 LVDDTVIGKLTAKINFYSLNETGLDGVPAFKEGVDLAG----GAIDEQFTYHYIKN 59
   :|| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 91 VVDSDDVRLVIARKEEFHAILEEBELKGAVVLIIFANKQDLFGALDDAAVTEALELHKIKN 150
   :|| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||

RESULT 6
US-10-739-930-10002
; Sequence 10002, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 10002
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) -- (188)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C235_53.p
US-10-739-930-10002
```

	Query Match	20.4%	Score 67	DB 5	Length 188
	Best Local Similarity	36.7%	Pred. No. 1.3		
	Matches 22	Conservative 7	Mismatches 25	Indels 6	Gaps 2
Qy	6	LVDDTVTIGKLTAKINFYSLNLTGLDQ--VPAPFKEGVDLAG-----GAIDQFTYHYKN	59		
	:		:		
	:		:		
	:		:		
Db	97	WVDSDDRLVTAKEEFHSILEEDELKAGVVLVTANKQDLPFGALDDAAITSELHKIKS	156		
	:		:		
	:		:		
	:		:		

```

RESULT 7
US-10-732-923-2799
; Sequence 2799, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(53796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 2799
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Asterina pectinifera
US-10-732-923-2799

```

```
Query Match      19.8%; Score 65; DB 5; Length 445;
Best Local Similarity 26.3%; Fred.No. 7.9;
Matches 15; Conservative 12; Mismatches 18; Indels 12; Gaps 1;
```

RESULT 8
US-10-282-122A-61809
; Sequence 61809, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 61809
; LENGTH: 601
; TYPE: PRT
; ORGANISM: Mycobacterium avium
US-10-289-122A-61809

```

```

Query Match      19.6%; Score 64.5; DB 4; Length 601;
Best Local Similarity 32.7%; Pred. No. 14;
Matches 17; Conservative 8; Mismatches 24; Indels 3; Gaps 1;

Qy      6 LVDDTVIGKLRATKIN---FYSLINETGLDGVPAKPGVDLAGGAIDQFTY 54
Db      257 LVADVPVGCILSGVDSLLVGLLAGAAGHGLSTFISPFESAGVGKDEPOY 308

```

```

RESULT 9
US-10-767-701-32620
; Sequence 32620, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Mol
; TITLE OF INVENTION: Plants and Uses Thereof For Plant I
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 32620
; LENGTH: 100
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-Cl4541_1.p
; US-10-767-701-32620

```

	Query Match	19.5%	Score 64	DB 4	Length 100
	Best Local Similarity	35.0%	Pred. No. 1.5		
	Matches 21	Conservative 8	Mismatches 25	Indels 6	Gaps 2
QY	6	LVDDTVTICKLRKTAKINFYSLNLTGDLG	---VPAFKEGVDLAG	---	GAIDQFTYHYIKN 59
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:
nb	9	WDSSDTRIVTAKEEHAIIEEDLGKGVLVYANKODLPFGALDDAAIT	ESLELHKIKS 68		
	:	:	:	:	:
	:	:	:	:	:
	:	:	:	:	:

```

RESULT 10
US-10-425-115-210704
; Sequence 210704, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 210704
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Zea mays

```

```

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_123759C.1.pcp
US-10-425-115-210704

```

```
Query Match      19.5%; Score 64; DB 4; Length 124;
Best Local Similarity 35.0%; Pred. No. 2;
Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;
```

```

RESULT 11
US-10-425-114-61645
; Sequence 61645, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 61645
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3079-008-A4_FLI pep
US-10-425-114-61645

```

```
Query Match      19.5%; Score 64; DB 4; Length 138;  
Best Local Similarity   35.0%; Pred. No. 2.3;  
Matches 21; Conservative    8; Mismatches 25; Indels    6; Gaps    2;
```

Qy 6 LVDDTVIGKLTAKINFYSLLNLTGLDQ---VPAPKEGVDLAG----GAIDEQFTYYIKN 59
Db 47 WVDSDDRLVTAKEEFLHLEEDLKGAVLVYANKQDFPGALDDAAITESLBHKIS 106

```

RESULT 12
US-10-739-930-8484
; Sequence 8484, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21 (53377) B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8484
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7449_6.p
US-10-739-930-8484

```

```
Query Match      19.5%; Score 64; DB 5; Length 148;
Best Local Similarity 35.0%; pred.No. 2.5;
Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;
```

Db 57 VDBSSDTRLVTAKEEFHALLFEDELKGAUVLVYANKQDLPGLDAAITSELELHKIKS 116

RESULT 13
US-10-437-963-187208
; Sequence 187208, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with Plant Growth and Development
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187208
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURES:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_83934C.1.pep
US-10-437-963-187208

```
Query Match      19.5%; Score 64; DB 4; Length 182;
Best Local Similarity 35.0%; Pred.No.3.;
Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;

Qy 6 LVDDTGWIGKLRATAKINFYSLLNLTGDLGG--VPAFKEGVLDAG---GAIDEQFTYHYIKN 59
   ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 91 VDPSSDTRLVTAKESFAHLEDELUKGAVLVYANKODLPGLADDAATIESLELHKKS 150
```

```

RESULT 14
US-10-425-114-48053
; Sequence 48053, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong .
; APPLICANT: Zhou, Yihua .
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48053
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3068-017-F7_FLI.pep
US-10-425-114-48053

```

```

Query Match      19.5%; Score 64; DB 4; Length 193;
Best Local Similarity 35.0%; Pred. No. 3.5;
Matches 21; Conservative 8; Mismatches 25; Indels 6; Gaps 2;

Qy 6 LVDDTWIGKLRITAKINFYSLNLTGDLG---VPAFKGQVDIAG---GAIDEQFTYHYIKN 59
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
102 VVPSDDTRLVLTAKESFHAILEDELEKGAVLVYANKKDDLPALGDAAAIYESLELHKIKS 161
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```


Search completed: January 24, 2006, 20:54:43
Job time : 12.4241 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 1.04031 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-18
 Perfect score: 329
 Sequence: 1 LAATPLVDDTVYIGKLR~~AKI~~.....GAIDQFTYHYIKNEAIIEL 65

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 entries

Database : Published Applications AA New:*

1:	/cn2/6/pdata/1/pubaa/US08	NEW	PUB	pep.*
2:	/cn2/6/pdata/1/pubaa/US06	NEW	PUB	pep.*
3:	/cn2/6/pdata/1/pubaa/US07	NEW	PUB	pep.*
4:	/cn2/6/pdata/1/pubaa/PCT	NEW	PUB	pep.*
5:	/cn2/6/pdata/1/pubaa/US05	NEW	PUB	pep.*
6:	/cn2/6/pdata/1/pubaa/US10	NEW	PUB	pep.*
7:	/cn2/6/pdata/1/pubaa/US11	NEW	PUB	pep.*
8:	/cn2/6/pdata/1/pubaa/US60	NEW	PUB	pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	61.5	18.7	302	6	US-10-793-626-2980	Sequence 2980, Ap
2	59.5	18.1	137	6	US-10-793-626-530	Sequence 530, App
3	59	17.9	227	6	US-10-510-386-60	Sequence 60, Appl
4	52	15.8	195	7	US-11-055-822-914	Sequence 914, App
5	52	15.8	314	6	US-10-485-517-230	Sequence 230, App
6	52	15.8	762	7	US-11-055-822-912	Sequence 912, App
7	51.5	15.7	959	6	US-10-467-962B-4	Sequence 4, Appli
8	51	15.5	333	6	US-10-467-657-6450	Sequence 6450, Ap
9	51	15.5	437	6	US-10-467-657-5526	Sequence 5526, Ap
10	50	15.2	168	7	US-11-108-172-198	Sequence 198, App
11	50	15.2	412	7	US-11-055-823-1086	Sequence 1086, Ap
12	50	15.2	520	6	US-10-878-556A-38	Sequence 38, Appl
13	49.5	15.0	91	7	US-11-073-347-89	Sequence 89, Appl
14	49.5	15.0	147	7	US-11-073-347-90	Sequence 90, Appl
15	49.5	15.0	308	6	US-10-467-657-4192	Sequence 4192, Ap
16	49.5	15.0	530	6	US-10-523-038-41	Sequence 41, Appl
17	49.5	15.0	738	6	US-10-523-038-39	Sequence 39, Appl
18	49	14.9	298	6	US-10-467-657-4362	Sequence 4362, Ap
19	49	14.9	447	7	US-11-024-959-286	Sequence 286, App
20	49	14.9	465	6	US-10-793-626-1676	Sequence 1676, Ap
21	49	14.9	622	7	US-11-212-443-185	Sequence 185, App
22	48.5	14.7	129	6	US-10-467-657-4354	Sequence 4354, Ap
23	48.5	14.7	205	7	US-11-134-795-27	Sequence 27, Appl
24	48.5	14.7	345	6	US-10-467-657-252	Sequence 252, App
25	48.5	14.7	345	6	US-10-467-657-3086	Sequence 3086, Ap

Sequence 8, Appl
Sequence 62, Appl
Sequence 86, Appl
Sequence 1484, Appl
Sequence 60, Appl
Sequence 78, Appl
Sequence 82, Appl
Sequence 80, Appl
Sequence 74, Appl
Sequence 42, Appl
Sequence 76, Appl
Sequence 94, Appl
Sequence 96, Appl
Sequence 256, Appl
Sequence 96, Appl
Sequence 4, Appl
Sequence 181, Appl
Sequence 13, Appl
Sequence 566, Appl
Sequence 992, Appl
Sequence 69, Appl

ALIGNMENTS

RESULT 1

```

US-10-793-626-2980
; Sequence 2980, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EP
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10793
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2980
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Ar
; OTHER INFORMATION: amino acid sequen
US-10-793-626-2980

```

Query Match 18.7%; Score 61.5; DB 6; Length 302;
Best Local Similarity 32.1%; Pred. No. 1.3;
Matches 18; Conservative 10; Mismatches 17; Indels 11; Gaps

RESIST 2

```

US-10-793-626-530
; Sequence 530, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU348005
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 530
; LENGTH: 137

```


Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

Qy 23 YSLNETGLDGVPAFKEGVLDAGAIDEQF 52
Db 332 YEVKELGLEDSFKFLAMELEQIAIKDPF 361

RESULT 10

US-11-108-172-198
; Sequence 198, Application US/11108172
; Publication No. US20050260177A1

; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun

; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather

; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy

; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong

; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.

; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun

; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.

; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.

; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.471C15
; CURRENT APPLICATION NUMBER: US/11/108,172

; PRIOR FILING DATE: 2005-04-15
; PRIOR FILING DATE: 2001-12-19

; PRIOR FILING DATE: 2001-08-03
; PRIOR FILING DATE: 2001-08-03

; PRIOR FILING DATE: 2001-04-10
; PRIOR FILING DATE: 2000-08-28

; PRIOR FILING DATE: 2000-06-29
; PRIOR FILING DATE: 2000-05-19

; PRIOR FILING DATE: 2000-03-06
; PRIOR FILING DATE: 2000-02-15

; PRIOR FILING DATE: 2000-01-10
; PRIOR FILING DATE: 2000-01-10

; PRIOR FILING DATE: 1999-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1130
; SEQ ID NO 198

; TYPE: PRT
; ORGANISM: Homo sapiens

US-11-108-172-198

Query Match 15.2%; Score 50; DB 7; Length 168;
Best Local Similarity 32.4%; Pred. No. 17;

Matches 11; Conservative 7; Mismatches 12; Indels 4; Gaps 1;

Qy 9 DTIVIGKLRITAKINFYSLNETGLDGVPAFKEGVD 42
Db 133 ETIIDKSKVKTNLQLFESGNWDI----EGID 162

RESULT 11

US-11-055-822-1086

Query Match 15.2%; Score 50; DB 6; Length 520;
Best Local Similarity 32.4%; Pred. No. 77;

Matches 11; Conservative 10; Mismatches 28; Indels 2; Gaps 2;

Qy 7 VDDTVIGKLRITAKINFYSLNETGLDGVPAFKEGVDLA-GGAIDEQFYHYIKNEAI 62
Db 201 IDDTPLDPERDYV-AASLYLQSGNEGMTALTRGTAPAQGTGIVDVQSTIGYLSNNV 256

RESULT 12
US-10-878-556A-38
; Sequence 38, Application US/10878556A
; Publication No. US20050266399A1

; GENERAL INFORMATION:
; APPLICANT: Hoffmann La-Roche Inc.
; TITLE OF INVENTION: HCV regulated protein expression

; FILE REFERENCE: 21762
; CURRENT APPLICATION NUMBER: US/10/878,556A
; CURRENT FILING DATE: 2004-06-28

; NUMBER OF SEQ ID NOS: 199
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38

; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens

; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: sw hum/hmcs_human
; DATABASE ENTRY DATE: 1993-07-01
US-10-878-556A-38

; Sequence 1086, Application US/11055822
; Publication No. US20050260707A1

; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus

; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig

; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor

; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CPCN

; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11

; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING DATE: 1999-06-25

; PRIOR FILING DATE: 1999-07-02
; PRIOR FILING DATE: 1999-08-12

; PRIOR FILING DATE: 2000-03-09
; PRIOR FILING DATE: 1999-07-01

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

; PRIOR FILING DATE: 1999-07-08
; PRIOR FILING DATE: 1999-07-08

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:21:03 ; Search time 82.6952 Seconds
(without alignments)
1966.133 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPODTISVLLDSRIQASRP.....AVSLVITQDIVYQNSLSA 370

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1889	100.0	370	ADU98761	Adu98761 Borrelia
2	1837	97.2	370	ADU98805	Adu98805 Borrelia
3	788	41.7	160	ADU98803	Adu98803 Borrelia
4	392	20.8	79	AAV83290	Aay83290 Borrelia
5	329	17.4	65	ADU98759	Adu98759 Borrelia
6	137.5	7.3	1633	ABu42513	Abu42513 Protein e
7	134.5	7.1	2295	3 AAB18180	ABu42513 Protein e
8	132.5	7.0	1155	3 AAG82343	ABu42513 Protein e
9	130.5	6.9	1802	3 AAY83170	Aay83170 Cell wall
10	130.5	6.9	1802	3 AAY70119	Aay70119 Staph. ep
11	128.5	6.8	1979	3 AAB18171	AAB18171 Plasmid
12	120.5	6.4	575	8 ADS08188	ADs08188 Staphyloc
13	120.5	6.4	598	6 ABU26292	ABu26292 Protein e
14	119	6.3	1178	6 ABU44244	ABu44244 Protein e
15	119	6.3	1331	6 ABU35503	ABu35503 Protein e
16	119	6.3	1411	2 AAY31948	Aay31948 Plasmid
17	117.5	6.2	5005	6 ABU48947	ABu48947 Protein e
18	117	6.2	1156	8 ADN17390	ADn17390 Bacterial
19	117	6.2	1417	2 AAY31947	Aay31947 Plasmid
20	117	6.2	1471	8 ADP25445	ADp25445 Plasmid
21	116.5	6.2	1191	6 ABU24124	ABu24124 Protein e
22	116	6.1	524	7 ADFO4502	ADf04502 Bacterial
23	115.5	6.1	439	7 ADI63080	ADI63080 Human apo
24	115.5	6.1	720	5 ABB77611	ABb77611 ANEPV RNA

25 115.5 6.1 1904 8 ADP25424 Adu98761 standard; protein; 370 AA.
26 115 6.1 590 6 ABU24573
27 115 6.1 4134 2 AAY31946
28 114.5 6.1 1100 3 AAB18213
29 113.5 6.0 596 6 ABU24511
30 113.5 6.0 1516 3 AAB18195
31 113.5 6.0 1781 8 ADP25447
32 113.5 6.0 3029 8 ADP25433
33 113 6.0 1346 3 AAB18236
34 113 6.0 1805 6 ABU35589
35 113 6.0 1805 7 ABO23572
36 112.5 6.0 326 5 ABU51677
37 112.5 6.0 382 2 AAW06932
38 112.5 6.0 1285 5 AABP7331
39 112 5.9 1254 2 AAW07503
40 112 5.9 1254 2 AAW24575
41 112 5.9 1881 5 AABP73809
42 111.5 5.9 1172 5 AABP73666
43 111 5.9 1087 2 AAY19935
44 111 5.9 1099 2 AAW40538
45 111 5.9 1119 2 AAY19934

ALIGNMENTS

RESULT 1
ADU98761
XX ADU98761 standard; protein; 370 AA.
AC ADU98761;
XX
DT 24-FEB-2005 (first entry)
XX
DE Borrelia burgdorferi antigenic polypeptide seqid 20.
XX
KW antibacterial; vaccine; immune stimulation; immunity; antigen;
KW DNA library.
XX
OS Borrelia burgdorferi.
XX
FN WO2004103269-A2.
XX
PD 02-DEC-2004.
XX
PF 17-OCT-2003; 2003WO-US033056.
XX
PR 18-OCT-2002; 2002US-0419401P.
XX
PA (MACR-) MACROGENICS INC.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Sykes KF, Hale KS, Johnston SA;
PI WPI; 2004-834155/82.
XX N-PSDB; ADU98760.
DR
PT Immunizing a subject against Borrelia burgdorferi infection comprises providing to the subject at least one Borrelia antigen or its fragment.
XX
XX Claim 27; SEQ ID NO 20; 121pp; English.
XX
XX The invention describes a method of immunizing a subject comprising providing to the subject at least one Borrelia antigen or its fragment to induce an immune response. Also described are: an isolated polynucleotide comprising a sequence having at least 17 contiguous nucleotides in common with a sequence not given in the specification; an isolated polypeptide having at least 5 consecutive amino acids of the sequence not given in the specification; a vaccine composition comprising at least one Borrelia antigen or at least one polynucleotide encoding a Borrelia antigen; screening for at least one test polypeptide or test polynucleotide encoding a polypeptide for an ability to produce an immune response; preparing a vaccine; vaccinating a subject; treating a subject infected

CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX
 SQ Sequence 370 AA;

Query Match 100.0%; Score 1889; DB 8; Length 370;
 Best Local Similarity 100.0%; Pred. No. 5.5e-154; Mismatches 0; Indels 0; Gaps 0;
 Matches 370; Conservative 0;

QY 1 MPQDTISVSLDSRIQASRPNNYPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
 DB 1 MPQDTISVSLDSRIQASRPNNYPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
 QY 61 EKXNGQDQFGKKTLLKTAMSNFNSSEESLSKADLFYKDKPEELKKYLKVRHTFV 120
 DB 61 EKXNGQDQFGKKTLLKTAMSNFNSSEESLSKADLFYKDKPEELKKYLKVRHTFV 120
 QY 121 VLINTEGNSDGLKIYKDDYDKFTPTSPFFVSTKEQEIKELFKDKGNTKERNIAVYS 180
 DB 121 VLINTEGNSDGLKIYKDDYDKFTPTSPFFVSTKEQEIKELFKDKGNTKERNIAVYS 180
 QY 181 NKNLNLHLKFIQYLHQASIFHAVNPGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
 DB 181 NKNLNLHLKFIQYLHQASIFHAVNPGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
 QY 241 LDGVPAPFKEGVDLAGGAIDQFTYHYKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
 DB 241 LDGVPAPFKEGVDLAGGAIDQFTYHYKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
 QY 301 AYTSAECLLKRFPVDRGLIIQYKNLSLTSPTPQLKLELSVNITNFNSINAVSLVITTDQ 360
 DB 301 AYTSAECLLKRFPVDRGLIIQYKNLSLTSPTPQLKLELSVNITNFNSINAVSLVITTDQ 360
 QY 361 IVDYQNSLSA 370
 DB 361 IVDYQNSLSA 370

RESULT 2
 ADU98805
 ID ADU98805 standard; protein; 370 AA.
 AC ADU98805;
 XX
 XX 24-FEB-2005 (first entry)
 XX
 XX Borrelia burgdorferi antigenic polypeptide seqid 64.
 XX
 XX antibacterial; vaccine; immune stimulation; immunity; antigen;
 KW DNA library.
 XX
 XX Borrelia burgdorferi.
 OS
 XX WO2004103269-A2.
 PN
 XX 02-DEC-2004.
 PD
 XX 17-OCT-2003; 2003WO-US033056.
 PF
 XX 18-OCT-2002; 2002US-0419401P.
 PR
 XX (MACR-) MACROGENICS INC.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 XX Sykes KF, Hale KS, Johnston SA;
 PI WPI; 2004-834155/82.
 XX
 XX N-PSDB; ADU98804.
 DR
 XX

PT Immunizing a subject against Borrelia burgdorferi infection comprises
 PT providing to the subject at least one Borrelia antigen or its fragment.
 XX
 XX Claim 27; SEQ ID NO 64; 121pp; English.

CC The invention describes a method of immunizing a subject comprising
 CC providing to the subject at least one Borrelia antigen or its fragment to
 CC induce an immune response. Also described are: an isolated polynucleotide
 CC comprising a sequence having at least 17 contiguous nucleotides in common
 CC with a sequence not given in the specification; an isolated polypeptide
 CC having at least 5 consecutive amino acids of the sequence not given in
 CC the specification; a vaccine composition comprising at least one Borrelia
 CC antigen or at least one polynucleotide encoding a Borrelia antigen;
 CC screening for at least one test polypeptide or test polynucleotide
 CC encoding a polypeptide for an ability to produce an immune response;
 CC preparing a vaccine; vaccinating a subject; treating a subject infected
 CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
 CC ; and detecting antibodies to a herpesvirus. The method is useful in
 CC immunizing a subject against Borrelia burgdorferi infection. This is the
 CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
 CC useful in immunizing a subject against infection by a member of the
 CC Borrelia species.
 XX

SQ Sequence 370 AA;

Query Match 97.2%; Score 1837; DB 8; Length 370;
 Best Local Similarity 97.0%; Pred. No. 1.7e-149; Indels 0; Gaps 0;
 Matches 359; Conservative 6; Mismatches 5;

QY 1 MPQDTISVSLDSRIQASRPNNYPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
 DB 1 MPQDTISVSLDSRIQASRPNNYPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
 QY 61 EKXNGQDQFGKKTLLKTAMSNFNSSEESLSKADLFYKDKPEELKKYLKVRHTFV 120
 DB 61 EKXNGQDQFGKKTLLKTAMSNFNSSEESLSKADLFYKDKPEELKKYLKVRHTFV 120
 QY 121 VLINTEGNSDGLKIYKDDYDKFTPTSPFFVSTKEQEIKELFKDKGNTKERNIAVYS 180
 DB 121 VLINTEGNSDGLKIYKDDYDKFTPTSPFFVSTKEQEIKELFKDKGNTKERNIAVYS 180
 QY 181 NKNLNLHLKFIQYLHQASIFHAVNPGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
 DB 181 NKNLNLHLKFIQYLHQASIFHAVNPGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
 QY 241 LDGVPAPFKEGVDLAGGAIDQFTYHYKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
 DB 241 LDGVPAPFKEGVDLAGGAIDQFTYHYKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
 QY 301 AYTSAECLLKRFPVDRGLIIQYKNLSLTSPTPQLKLELSVNITNFNSINAVSLVITTDQ 360
 DB 301 AYTSAECLLKRFPVDRGLIIQYKNLSLTSPTPQLKLELSVNITNFNSINAVSLVITTDQ 360
 QY 361 IVDYQNSLSA 370
 DB 361 IVDYQNSLSA 370

RESULT 3
 ADU98803
 ID ADU98803 standard; protein; 160 AA.
 XX
 AC ADU98803;
 XX
 XX 24-FEB-2005 (first entry)
 DT
 XX Borrelia burgdorferi antigenic polypeptide seqid 62.
 DE
 XX antibacterial; vaccine; immune stimulation; immunity; antigen;
 KW DNA library.
 XX
 XX Borrelia burgdorferi.
 OS

PN WO2004103269-A2.
XX 02-DEC-2004.
XX 17-OCT-2003; 2003WO-US033056.
XX 18-OCT-2002; 2002US-0419401P.
XX (MACR-) MACROGENICS INC.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Sykes KF, Hale KS, Johnston SA;
XX WPI; 2004-834155/82.
XX N-PSDB; ADU98802.
XX Immunizing a subject against Borrelia burgdorferi infection comprises
PT providing to the subject at least one Borrelia antigen or its fragment.
XX Claim 27; SEQ ID NO 62; 121pp; English.
XX The invention describes a method of immunizing a subject comprising
CC providing to the subject at least one Borrelia antigen or its fragment to
CC induce an immune response. Also described are: an isolated polynucleotide
CC comprising a sequence having at least 17 contiguous nucleotides in common
CC with a sequence not given in the specification; an isolated polypeptide
CC having at least 5 consecutive amino acids of the sequence not given in
CC the specification; a vaccine composition comprising at least one Borrelia
CC antigen or at least one polynucleotide encoding a Borrelia antigen;
CC screening for at least one test polypeptide or test polynucleotide
CC encoding a polypeptide for an ability to produce an immune response;
CC preparing a vaccine; vaccinating a subject; treating a subject infected
CC with a pathogen; passive immunization; vaccination; detecting Herpesvirus
CC ; and detecting antibodies to a herpesvirus. The method is useful in
CC immunizing a subject against Borrelia burgdorferi infection. This is the
CC amino acids sequence of a Borrelia burgdorferi polypeptide that may be
CC useful in immunizing a subject against infection by a member of the
CC Borrelia species.
XX Sequence 160 AA;
SQ
Query Match 41.7%; Score 788; DB 8; Length 160;
Best Local Similarity 98.1%; Pred. No. 1.3e-59;
Matches 157; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 211 LAATPLVDDTVIGKLTAKINFYLLNETGLDGVPAFKGVDLGAGIDEQFTYHYKNE 270
DB 1 LAATPLVDDTVIGKLTAKINFYLLNETGLDGVPAFKGVDLGAGIDEQFTYHYKNE 60
QY 271 AIIEILIRIWNKNRQNSKLSALQSGARDNAYTSAIECLLKRFVDRGLIIQYKNLSLTLS 330
DB 61 AIIEILIRIWNKNRQNSKLSALQSGARDNAYTSAIECLLKRFVDRGLIIQYKNLSLTLS 120
QY 331 PTPQLKLELSVNTYFNFSINAVSLVTTQDIVDYQNSLSA 370
DB 121 PTPQLKLELSVNTYFNFSINAVSLVTTQDIVDYQNSLSA 160
RESULT 4
AA83290
ID AAY83290 standard; protein; 79 AA.
XX AAY83290;
XX 12-SEP-2003 (revised)
DT 16-AUG-2000 (first entry)
XX Borrelia antigenic polypeptide.
XX Diagnosis; detection; immune response; antigen; antibody; immunogen;
XX vaccine; treatment; Lyme borelliosis.
XX Borrelia burgdorferi; ZS7.

XX WO200022134-A2.
PN 20-APR-2000.
XX 12-OCT-1999; 99WO-EP007651.
XX 13-OCT-1998; 98DE-01047142.
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX Simon M, Wallich R, Kramer M;
XX WPI; 2000-329166/28.
XX N-PSDB; AAZ93798.
XX Diagnosing borelliosis by detecting specific antibodies, using new
PT Borrelia antigens expressed preferentially during infection and also
PT useful as immunogens in vaccines.
XX Claim 11; Page 44; 56pp; German.
XX Diagnosis of borelliosis infection can be achieved by detecting anti-
CC Borrelia antibodies. The antibodies are detected using a detection
CC reagent that contains many Borrelia antigens and reacting this reagent
CC with serum from Borrelia infected animals. The antigens have molecular
CC weights of about 9.5, 18, 19, 30, 32, 33, 62, 70, 80, 90, 100 and/or 102
CC kD. The method is particularly used to diagnose Lyme borelliosis. The
CC antigens (or cell lysate or fractions containing them or nucleic acid
CC encoding them) are also useful as immunogens for production of
CC antibodies. They can also be used in vaccines for treatment and
CC prevention of borelliosis. This method of detection is advantageous since
CC it detects immunological structures in Borrelia that are expressed
CC selectively in vivo and provides a more reliable diagnosis than existing
CC methods. See GENESEQ records AAZ93790-293807 and AAY83282-183299.
XX (Updated on 12-SEP-2003 to standardise OS field)
XX Sequence 79 AA;
SQ
Query Match 20.8%; Score 392; DB 3; Length 79;
Best Local Similarity 94.9%; Pred. No. 6.7e-26;
Matches 75; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 66 NGQDFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHRTFVLLNT 125
DB 1 NGQDFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHRTFVLLNT 60
QY 126 EGDNSDDGLKIYKDDYDKF 144
DB 61 EGDASDDGLKIYKODYNKF 79
RESULT 5
ADU98759
ID ADU98759 standard; protein; 65 AA.
XX AC
XX ADU98759;
XX 24-FEB-2005 (first entry)
XX Borrelia burgdorferi antigenic polypeptide seqid 18.
XX antibacterial; vaccine; immune stimulation; immunity; antigen;
XX DNA library.
XX Borrelia burgdorferi.
XX WO2004103269-A2.
XX 02-DEC-2004.
XX 17-OCT-2003; 2003WO-US033056.

PR	18-OCT-2002; 2002US-0419401P.	XX	PA	(ELIT-) ELITRA PHARM INC.	XX
XX	(MACR-) MACROGENICS INC.	XX	PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;	
PA	(TEXA) UNIV TEXAS SYSTEM.	PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;		
XX		XX	WPI; 2003-029926/02.		
PI	Sykes KF, Hale KS, Johnston SA;	DR	N-PSDB; ACN46383.		
XX		DR			
DR	WPI; 2004-834155/82.	XX	New antisense nucleic acids, useful for identifying proteins or screening		
XX	N-PSDB; ADU98758.	PT	for homologous nucleic acids required for cellular proliferation to		
PT	Immunizing a subject against Borrelia burgdorferi infection comprises	PT	isolate candidate molecules for rational drug discovery programs.		
PT	providing to the subject at least one Borrelia antigen or its fragment.	XX			
XX		XX	Claim 27; SEQ ID NO 18; 121pp; English.		
PS		PS	The invention describes a method of immunizing a subject comprising		
XX		CC	providing to the subject at least one Borrelia antigen or its fragment to		
CC		CC	induce an immune response. Also described are: an isolated polynucleotide		
CC		CC	comprising a sequence having at least 17 contiguous nucleotides in common		
CC		CC	with a sequence not given in the specification; an isolated polypeptide		
CC		CC	having at least 5 consecutive amino acids of the sequence not given in		
CC		CC	the specification; a vaccine composition comprising at least one Borrelia		
CC		CC	antigen or at least one polynucleotide encoding a Borrelia antigen;		
CC		CC	screening for at least one test polypeptide or test polynucleotide		
CC		CC	encoding a polypeptide for an ability to produce an immune response;		
CC		CC	preparing a vaccine; vaccinating a subject; treating a subject infected		
CC		CC	with a pathogen; passive immunization; vaccination; detecting Herpesvirus		
CC		CC	; and detecting antibodies to a herpesvirus. The method is useful in		
CC		CC	immunizing a subject against Borrelia burgdorferi infection. This is the		
CC		CC	amino acids sequence of a Borrelia burgdorferi polypeptide that may be		
CC		CC	useful in immunizing a subject against infection by a member of the		
CC		CC	Borrelia species.		
XX		XX			
SQ	Sequence 65 AA;	SQ			
			Query Match		
			Best Local Similarity 17.4%; Score 329; DB 8; Length 65;		
			Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	211 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKGVLAGGAIHQFTYHYKNE 270	QY	18 SRPNYYNPLLVYKTAIKVKNK-DAANYKILNLTVNNYKQIETLEKNGNGQDQFGKEKT 76		
Db	1 LAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKGVLAGGAIHQFTYHYKNE 60	Db	113 SKPKANEAVVTNESTKPKTEAPTNEESIAETPKTSTTQDDSTKKNPFLKDLNLSST 172		
QY	271 AIEL 275	QY	77 LLKTAMSNFFNSSESLKADLFYK-----DKPEELKKYLKVRHHTFVLINTEGDS 130		
Db	61 AIEL 65	Db	173 TSKE-----KTDEHSTKQAMSTNKNLDTNDSPTQSEK-----TSSQANNDSTNQ 220		
		QY	131 D-----DGLKIYKDDYD-----KFKTPSIFVFVSTKEQE-----LKE 162		
		Db	221 SAPSKQLDSKPSKQVYKTKFNDEPTQDVETTYTKLTPSITDSSVNDKQDYTSASAS 280		
		QY	163 LFKDKGNTEKERNIAVYSNNKDLHLKFIIS-QYLHOASIFHAV-----NP-YGM--PLAA 213		
		Db	281 LGVDSNETE-----AITNAVRNLDLKTASREQINEAIIAEALKKDFSNPDYGVDTPLAL 335		
		QY	214 TPLVDDTVIGKLTAKINFYSLNLTG-----DGVPAKKEGVDL-----AGGAI----- 258		
		Db	336 NTSQSKTSPHKASPRMNLMSLAAPNSGKNVNDKVTNPTLSLNKNNHANNVWPTS 395		
			Query Match		
			Best Local Similarity 7.3%; Score 137.5; DB 6; Length 1633;		
			Matches 97; Conservative 58; Mismatches 168; Indels 107; Gaps 21;		
RESULT 6					
ABU42513					
ID	ABU42513 standard; protein; 1633 AA.				
XX					
AC	ABU42513;				
XX					
DT	19-JUN-2003 (first entry)				
XX	Protein encoded by Prokaryotic essential gene #28040.				
DE					
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.				
KW					
XX	Staphylococcus epidermidis.				
OS					
XX	WO200277183-A2.				
PN					
XX	03-OCT-2002.				
PD					
XX	21-MAR-2002; 2002WO-US009107.				
XX					
PR	21-MAR-2001; 2001US-00815242.				
PR	06-SEP-2001; 2001US-00948993.				
PR	25-OCT-2001; 2001US-0342923P.				
PR	08-FEB-2002; 2002US-00072851.				
PR	06-MAR-2002; 2002US-0362699P.				

QY 259 DEQFTY--HYIKNEAI-----IELIRIWNKNNRQNSKLSALQLSGAR 298
 DB 396 NEQFNKANYELDDSIKEDGFTTIKYQYIRPGGLEPAI---KTQLRSKDGSIYANGVY 452
 QY 299 DNAYTSAIECLLKRFPVDRGLIIQYKNL--SUTLSPTQLKLELSVNTYFNFINAVSLVI 356
 DB 453 DKT-TNTTYYTFTNYVD-----QYQNTGTFDLIATPKRETAIKDNQNPMEVTIANEW 506
 QY 357 TTQDIVDYON 366
 DB 507 KKOFIVDYG 516

RESULT 7

ID AAB18180 standard; protein; 2295 AA.

AC AAB18180;

DT 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:37.

DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

OS WO200025728-A2.

XX 11-MAY-2000.

XX 05-NOV-1999; 99US-05026796.

XX 05-NOV-1998; 98US-0107131P.

XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection.

XX Disclosure; Page 92-98; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (i) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are rifins or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification

XX

SQ Sequence 2295 AA;
 Query Match 7.1%; Score 134.5; DB 3; Length 2295;
 Best Local Similarity 21.5%; Pred. No. 0.13;
 Matches 95; Conservative 57; Mismatches 124; Indels 165; Gaps 23;
 QY 23 YNELLVYTKAKIKVNDAAANYKILNLTNNYKQIE-TLEKDNCGGQDFGKEKTLAKTA 81
 DB 648 YIPLIFNPTKDVFTSQINIKSNFNTIYELITITIKYKKNFYDD-----LUKCR 701
 QY 82 MSNFFNSSESLKSADLFYKDKPELKK-----YLVKVRHTFVVLINTEGNSDDGLKIY 137
 DB 702 REIFF-KDRHLENNIM---DKQBELKKIRNLMRIHE-----VSNEGNNRTINRKY 751
 QY 138 KD-----DYDKFKTPIFFVFSTEQEIKEI-----FKDKGNTEKERN-----TAVY 179
 DB 752 KKYGYDYDKNN--ELYVV-----EKNILNVNDTTFNFMNKEKDKNVFDINKTMRIYDY 805
 QY 180 SNN-----KDNLH--LK-----FISQYLHOA-----SIFHAVN----- 205
 DB 806 YNNINLNIPTAAIKMKDKIYDQLKLSNFVEKLNESICVLSFLYLIGINDNGKLHF 865
 QY 206 PYGMPLAATPLVDDTVIGKLTAKINFYSLNLTGDLGVPAFKGV----- 251
 DB 866 PYGFP-----RNIDFSVKLIREGKDLGNLFLSGVLVHINLPIFVNSS 908
 QY 252 -----DLAGGAIDEQFTYHYIKNEAI-----IELIRIWNKNNRQNSKLSAL 292
 DB 909 ISISTEMDDVLEWNNDSINSFFIYYKNNENIRNHDPLSDENRIIPKEDNIKSIIISY 968
 QY 293 QLSGARDNAYTSAIECLLKRFPVDRGLIIQYKNLSLTLSPTQLKLELSVNTY-----NF 347
 DB 969 SLGSKODFFSKL-----AFTNNVIRLKYKN-----KTNNYTLKDYDFD 1007
 QY 348 SINAVSLVITTDIVDYONSL 368
 DB 1008 -----TFDKINYKNSV 1018

RESULT 8

AAG82343

ID AAG82343 standard; protein; 1155 AA.

XX AAG82343;

XX 03-SEP-2001 (first entry)

XX S. epidermidis open reading frame protein sequence SEQ ID NO:1780.

XX Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
 KW endocarditis.

XX Staphylococcus epidermidis.

XX WO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000WO-US030782.

XX 09-NOV-1999; 99US-0164258P.

XX (GLAX) GLAXO GROUP LTD.

XX Kimmerly WJ;

XX WPI; 2001-316495/33.

XX N-PSDB; AAB53193.

XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis.

XX Claim 18; Page 491-492; 2188pp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
 CC and (II) can have antibacterial activity and therefore can be used in
 CC vaccination. The nucleic acids (I) may be used to produce the S.
 CC epidermidis polypeptides (II) via the production of vectors containing
 CC them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
 CC represent oligonucleotide sequences and primers which are used in the
 CC exemplification of the present invention. N.B. The present invention
 CC specifically claims all the polynucleotide sequences given in the
 CC sequence listing of the present specification, however the sequence
 CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
 CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 CC for SEQ ID NO:4455 to 4464
 XX
 SQ Sequence 1155 AA;

Query Match 7.0%; Score 132.5; DB 4; Length 1155;
 Best Local Similarity 22.3%; Pred. No. 0.072;
 Matches 96; Conservative 58; Mismatches 169; Indels 107; Gaps 21;
 QY 18 SRPNYNNPLLVYTKAKIKVKNK-DAANYKILNLTVNNYEKQIETLEKNGNGQDQFGKEKT 76
 DB 113 SKPKANEAVVTNESTKPKTTEAPTVEESIAETPKTSTTQQDSTKKNPSLKNLNSST 172
 QY 77 LLKTAMSNFFNSSESLKSADLFYIK-----DKPEELKKYLKVRHHTFVVLINTEGNS 130
 DB 173 TSKESS-----KTDEHSTKQAKMSTNKNLDTNDSPTQSEK-----TSSQANNDSTNQ 220
 QY 131 D-----DGLKIYKDDYD-----KFKTPSIFFFVSTKEQE-----IKE 162
 DB 221 SAPSKQLDSKPEQKVKYTKFNDPTQDVEHTTTKLTPTSISTDSSVNDKQDYTSASVAS 280
 QY 163 LFKDKGNTKERNIAVYNNKNDLHLKPFIS-QYLHQASIFHAV-----NP-YGM--PLAA 213
 DB 281 LGVDSNETE-----AITNAVRNDLKAASREQINEAIIAEALKKDFSNPDYGVDTPLAL 335
 QY 214 TPLVDDTVIGKLTAKINPYSILNTEGL-----DGVPAKFGVDI-----AGGAI----- 258
 DB 336 NTSQSKNSPHKASFPASPNMLMSLAEPNSGKVNVDKVTNPTSLNKNHANNVWPTS 395
 QY 259 DRQFTY--HYIKNEAI-----IELIRIWNKNRQNSKLSALQLSGAR 298
 DB 396 NEQFNLKANYELDDSIKEGDTFTIKYGVYIRPGGLELPAI---KTQLRSKQGSIVANGVY 452
 QY 299 DNAYTSAIECLKRFVDRGLLIIOYKNL--SLTSPPTQKLKLSVNITNFSINAVSLVI 356
 DB 453 DKT-TNTTYYTNTYVD-----QYQNIITGSPDLIATPKRETAIKDNQNPMEVTIANEVY 506
 QY 357 TTQDVIDVYON 366
 DB 507 KKDIFVDYGN 516

RESULT 9
 AAY83170
 ID AAY83170 standard; protein; 1802 AA.
 XX
 AC AAY83170;
 XX
 DT 24-JUL-2000 (first entry)
 XX
 DE Cell wall protein SdrF.
 XX
 KW SdrF; SdrG; SdrH; coagulase negative; staphylococcus; septicemia;
 KW osteomyelitis; endocarditis; immune response; vaccine; graft; stent;

KW intravenous catheter; heart valve; cardiac.
 XX Staphylococcus sp.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 12 /note= "Position encoded by TAA stop codon"
 FT Misc-difference 28 /note= "Position encoded by TAA stop codon"
 FT Misc-difference 1771 /note= "Position encoded by TAG stop codon"
 FT Misc-difference 1774 /note= "Position encoded by TGA stop codon"
 FT Misc-difference 1793 /note= "Position encoded by TAA stop codon"
 FT Misc-difference 1800 /note= "Position encoded by TGA stop codon"
 XX
 PN WO200012689-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 31-AUG-1999; 99WO-US019728.
 XX
 PR 31-AUG-1998; 98US-0098443P.
 PR 25-JAN-1999; 99US-0117119P.
 XX
 PA (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 XX
 PI Foster TJ, Hook M, Davis S, Hartford O, McCreia K, Ni Eidhin D;
 DR WPI; 2000-256637/22.
 DR N-PSDB; AAZ93533.
 XX
 PT Recombinant or synthetic proteins from coagulase-negative staphylococci
 PT useful for prevention, treatment and diagnosis of staphylococcal
 PT infections bind soluble and immobilized fibrinogen.
 XX
 PS Claim 2; Fig 2; 104pp; English.
 XX
 CC Isolated Staphylococcus Sdr cell wall proteins which bind both soluble
 CC and immobilized fibrinogen are useful for treating or preventing
 CC coagulase-negative staphylococcal infection such as septicemia,
 CC osteomyelitis or endocarditis, and for inducing immune responses in
 CC patients. The cell wall proteins are also useful for reducing coagulase-
 CC negative staphylococci of indwelling medical devices such as
 CC vascular grafts, vascular stents, intravenous catheters, artificial heart
 CC valves and cardiac assist devices. The cell wall associated proteins are
 CC able to inhibit staphylococcal adhesion to immobilised extracellular
 CC matrix or host cells present on the surface of implanted biomaterials
 XX
 SQ Sequence 1802 AA;

Query Match 6.9%; Score 130.5; DB 3; Length 1802;
 Best Local Similarity 22.1%; Pred. No. 0.2;
 Matches 95; Conservative 59; Mismatches 169; Indels 107; Gaps 21;
 QY 18 SRPNYNNPLLVYTKAKIKVKNK-DAANYKILNLTVNNYEKQIETLEKNGNGQDQFGKEKT 76
 DB 150 SKPKANEAVVTNESTKPKTTEAPTVEESIAETPKTSTTQQDSTKKNPSLKNLNSST 209
 QY 77 LLKTAMSNFFNSSESLKSADLFYIK-----DKPEELKKYLKVRHHTFVVLINTEGNS 130
 DB 210 TSKESS-----KTDEHSTKQAKMSTNKNLDTNDSPTQSEK-----TSSQANNDSTNQ 257
 QY 131 D-----DGLKIYKDDYD-----KFKTPSIFFFVSTKEQE-----IKE 162
 DB 258 SAPSKQLDSKPEQKVKYTKFNDPTQDVEHTTTKLTPTSISTDSSVNDKQDYTSASVAS 317
 QY 163 LFKDKGNTKERNIAVYNNKNDLHLKPFIS-QYLHQASIFHAV-----NP-YGM--PLAA 213
 DB 318 LGVDSNETE-----AITNAVRNDLKAASREQINEAIIAEALKKDFSNPDYGVDTPLAL 372

QY 214 TPLVDDTVIGKLTAKINFYLLNETGL-----DGVPAFKGVLDL-----AGGAI----- 258
 Db 373 NRSQSKNSPKHSASPRMNLMSLAAPNSGKNVNDVKITNPTLSLNKNHANNVIMPTS 432
 QY 259 DEQFTY--HYIKNEAI-----IELIRIWNKNRQNSKLSALQSGAR 298
 Db 433 NEQFNKANYELDDSIKEGDTFTIKYQYIRPGGLELPAI---KTQLRSKDGSIYANGVY 489
 QY 299 DNAYTSAIECLLKRFPVDRGLIIQYKNL--SLTSPPTQLKLELSVNTYFNFSINAVSLVI 356
 Db 490 DKT-TNTTYYFTNVVD-----QYQNTGSPDLIATPKRETAIKDNQNPMEVTIANEVV 543
 QY 357 TTQDIDVYQN 366
 Db 544 KKDFIVDYGN 553

RESULT 10
 AAY70119
 ID AAY70119 standard; protein; 1802 AA.
 AC AAY70119;
 XX
 DT 06-JUN-2000 (first entry)
 XX
 DE Staph. epidermidis serine-aspartate repeat region protein SdrF.
 XX
 KW Multicomponent vaccine; immunostimulatory; antibacterial; MSCRAMM;
 KW microbial surface components recognising adhesive matrix molecules;
 KW collagen binding protein; CBP; CNA; fibrinogen binding protein;
 KW Clumping factor A; ClfA; Clumping factor B; ClfB; FnBP;
 KW fibronectin binding protein; Staphylococcus infection;
 KW serine-aspartate repeat region protein; SDR protein; SdrF.
 XX
 OS Staphylococcus epidermidis.
 XX

Key Location/Qualifiers
 FT Misc-difference 12 /note= "Encoded by in-frame stop codon TAA"
 FT Misc-difference 28 /note= "Encoded by in-frame stop codon TAA"
 FT Misc-difference 1771 /note= "Encoded by in-frame stop codon TAG"
 FT Misc-difference 1774 /note= "Encoded by in-frame stop codon TGA"
 FT Misc-difference 1793 /note= "Encoded by in-frame stop codon TAA"
 FT Misc-difference 1800 /note= "Encoded by in-frame stop codon TGA"
 FT
 PN WO200012131-A1.
 XX
 PD 09-MAR-2000.
 XX
 PF 31-AUG-1999; 99WO-US019727.
 XX
 PR 31-AUG-1998; 98US-0098439P.
 XX
 PA (INH-) INHIBITEX INC.
 PA (TEXA) UNIV TEXAS A & M SYSTEM.
 PA (QUEB-) QUEEN ELIZABETH COLLEGE DUBLIN.
 XX
 PI Patti JM, Foster TJ, Hook M;
 XX
 DR WPI; 2000-237781/20.
 DR N-PSDB; AA251201.
 XX
 XX Composition used for generating immune response or for inhibiting
 PT microbial colonization in an animal comprises antibodies that bind
 PT collagen binding protein, fibrinogen binding protein and, optionally,
 PT fibronectin binding protein.
 XX

PS Claim 8; Fig 3; 115pp; English.
 XX The patent discloses multicomponent vaccines containing selected combinations of bacterial binding proteins termed MSCRAMM (microbial surface components recognising adhesive matrix molecules) or their antibodies. A vaccine composition is provided that includes collagen binding protein or peptide, e.g. CNA, a fibrinogen binding protein, preferably Clumping factor A (ClfA) or Clumping factor B (ClfB), and optionally a fibronectin binding protein e.g. FnBP-A. The vaccines are useful for imparting protection against a broad spectrum of Staphylococcal strains and for inhibiting microbial colonisation, especially of Staphylococcus aureus, in an animal. The combinations can also be used to select donor blood pools for the preparation of purified blood products for passive immunisation. The present sequence is a serine-aspartate repeat region protein, SdrF from Staphylococcus epidermidis. The Sdr protein is useful in vaccine preparation in combination with specific bacterial binding proteins. These vaccines can be used to treat a broad spectrum of bacterial infections, including those arising from both coagulase-positive and coagulase-negative bacteria

QY 18 SRPNYPLLVYTKAKIVNK-DAANYKILNLTNNYEKQIETLEKNGQDQFGKEKT 76
 Db 150 SKPKANEAVVNESTKPKTTEAPTVEESIAETPKTSTQDSTEKNPNLKNLSST 209
 QY 77 LLKTAMSNFFNSSESLSADLFYK-----DKPEELKYLKVRHRTFVVLINTEGDSN 130
 Db 210 TSKE-----KTDEHSTKQAMSTKNSLNDTNDSTQSEK-----TSSQANNDSTNQ 257
 QY 131 D-----DGLKIYKDDYD-----KFKTPSIFVFSTKEQE-----IKE 162
 Db 258 SAPSKQLDSKPEQKVKYKTKENDEPTQDVEHTTTKLTSPSVSTDSSVNDKQDYTRSAVAS 317
 QY 163 LFKQKNGTEKERNIAVSNKNDNLHLKFIIS-OYLHOASIFHAV-----NP-YGM--PLAA 213
 Db 318 LGVDSNETE-----AITNAVRDNLDLKAASREQINEAIIAEALKKDFSNPDYVDTPAL 372
 QY 214 TPLVDDTVIGKLTAKINFYLLNETGL-----DGVPAFKGVLDL-----AGGAI----- 258
 Db 373 NRSQSKNSPKHSASPRMNLMSLAAPNSGKNVNDVKITNPTLSLNKNHANNVIMPTS 432
 QY 259 DEQFTY--HYIKNEAI-----IELIRIWNKNRQNSKLSALQSGAR 298
 Db 433 NEQFNKANYELDDSIKEGDTFTIKYQYIRPGGLELPAI---KTQLRSKDGSIYANGVY 489
 QY 299 DNAYTSAIECLLKRFPVDRGLIIQYKNL--SLTSPPTQLKLELSVNTYFNFSINAVSLVI 356
 Db 490 DKT-TNTTYYFTNVVD-----QYQNTGSPDLIATPKRETAIKDNQNPMEVTIANEVV 543
 QY 357 TTQDIDVYQN 366
 Db 544 KKDFIVDYGN 553

Query Match 6.9%; Score 130.5; DB 3; Length 1802;
 Best Local Similarity 22.1%; Pred. No. 0.2;
 Matches 95; Conservative 59; Mismatches 169; Indels 107; Gaps 21;
 SQ Sequence 1802 AA;

RESULT 11
 AAB18171
 ID AAB18171 standard; protein; 1979 AA.
 XX
 AC AAB18171;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:28.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.

XX WO200025728-A2.
XX 11-MAY-2000.
XX 05-NOV-1999; 99WO-US026796.
XX 05-NOV-1998; 98US-0107131P.
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection.
XX
XX Disclosure; Page 70-75; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II). (I) and
XX (II) are useful for the development of vaccines against P. falciparum
XX infection. (I) and polyclonal antisera or a monoclonal antibody raised to
XX immunogens comprising the sequences of (I), are useful in the detection
XX of infection with P. falciparum. Furthermore, (I) (especially when they
XX are refined or secreted or membrane proteins) can aid the identification
XX of drugs to treat or prevent P. falciparum infection, or they can be used
XX to identify drug resistance in P. falciparum. Sequencing of the
XX Plasmodium chromosome 2 and the subsequent identification of proteins
XX encoded by it will help to expand our understanding of parasite biology,
XX a process hampered by the complexity of the parasitic lifecycle, and
XX provide new targets for vaccine and drug development. Parasite resistance
XX to drugs and mosquito resistance to insecticides have led to a resurgence
XX of malaria in many parts of the world, and there is a pressing need for
XX vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
XX represent nucleotide and protein sequences given in the present
XX invention, but which are not specifically mentioned within the
XX specification
XX
XX Sequence 1979 AA;

Query Match
Best Local Similarity 6.8%; Score 128.5; DB 3; Length 1979;
Matches 77; Conservative 57; Mismatches 128; Indels 107; Gaps 16;

QY 3 QDTISVLLDSRIQASR-----PNYNPL-----VYTKAK--IKNKDAAN-- 42
DB 242 RETISIEIDIKKHLEKLEKIEKEKEDLENLKKLLSKENYLBKLGCVKEKNTINSL 301
QY 43 -----YKILNLTVNNYKEQIETLEKONGGQDQFGKEKTLTKTAMSNFFNSSES 92
DB 302 NNDNIETEKKYKLLYELEKKNQIDLLNKQBEKEKEKEKEKEK-----BREKEK 352
QY 93 LKSADLFYKPKPEELKYLKVH-----RHTFVVLINTEGD-----NSDDGL 134
DB 353 EKEYDTLIKELKDEKISILEKVSIKVREMDIEKREHNFLEHEDQLKDKNSFVKNQQL 412
QY 135 KLYKDDYDKFKTPSPFFVSTKEQELKEL-----FKDKGN-----TEKERNIAVYS--NNKD 184
DB 413 KYVKEIENLKT-----LEKKEKELKDTENVSKEEINKLINQLNKEKEKQILAFKNHKE 467
QY 185 NLH-LKFTISQYLHQASIFHAVNPVPGMPLAATPLVDVTGKLTAKINFYSLNLTGDD 243
DB 468 EHLGLK-----EELKESVKITKETIQLQELQWMDIKQKELDQ 503
QY 244 VPAFKEGVDLAGGAID---EQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300

DB 504 LQE-KYNAQIESISIELSKKEYNQYKNTYIEEINNINLNKLEETNKYTNLIQ-----N 556
QY 301 AYTSALIECL 309
DB 557 NYTNEINML 565
RESULT 12
ADS08188
ID ADS08188 standard; protein; 575 AA.
XX AC ADS08188;
XX 04-NOV-2004 (first entry)
XX Staphylococcus epidermis polypeptide seqid 7483.
XX
XX antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
XX recombinant expression vector; infection; computer readable medium;
XX computer based system.
XX
XX Staphylococcus epidermidis.
XX
XX US2004147734-A1.
XX 29-JUL-2004.
XX 01-DEC-2003; 2003US-00724972.
XX 08-NOV-1997; 97US-0064964P.
XX 13-AUG-1998; 98US-00134001.
XX 29-NOV-1999; 99US-00450969.
XX (DOUC/) DOUCETTE-STAMM L.
XX (BUSH/) BUSH D.
XX Doucette-Stamm L, Bush D;
XX WPI; 2004-580138/56.
XX N-PSDB; ADS04416.
XX
XX New isolated polypeptide and encoding nucleic acid derived from
XX Staphylococcus epidermidis, useful for diagnosing, preventing and/or
XX treating an S. epidermidis bacterial infection.
XX
XX Claim 17; SEQ ID NO 7483; 741pp; English.
XX
XX The invention describes an isolated nucleic acid comprising a nucleotide
XX sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
XX 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
XX of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
XX given in the specification. Also described are: a recombinant expression
XX vector; a cell comprising a recombinant expression vector of (I);
XX producing an S. epidermidis polypeptide; an isolated nucleic acid
XX comprising a nucleotide sequence of at least 8 nucleotides in length; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection, comprising a nucleic acid cited above and a carrier; treating
XX a subject for S. epidermidis infection; a recombinant or substantially
XX pure preparation of an S. epidermidis polypeptide or its fragment; a
XX vaccine composition for prevention or treatment of an S. epidermidis
XX infection; detecting the presence of a Staphylococcus nucleic acid in a
XX sample; a computer readable medium having recorded in it the nucleotide
XX sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
XX system for identifying fragments of the Staphylococcus genome of
XX commercial importance; a computer based system for identifying fragments
XX of the Staphylococcus plasmids of commercial importance; identifying
XX commercially important nucleic acid fragments of the Staphylococcus
XX genome and/or plasmids; and identifying an expression modulating fragment
XX of the Staphylococcus genome and/or plasmids. The methods and
XX compositions of the present invention are useful for the diagnosis,
XX prevention and/or treatment of an Staphylococcus epidermidis bacterial
XX infection. This is the amino acid sequence of a S. epidermidis protein of


```
CC the invention.
XX
SQ Sequence 575 AA;

Query Match      6.4%; Score 120.5; DB 8; Length 575;
Best Local Similarity 20.6%; Pred. No. 0.29;
Matches 84; Conservative 68; Mismatches 131; Indels 125; Gaps 22;

QY 29 YKTAKIKVNDK-ANYKILNLTNNYKQIETLEKDNQ-GDQPGKEKTLTKTAMS--- 83
DQ 69 YKAWKNEFDROVQKGAQKSMIIPIVQKRNVEIKDDNGNIVRDNNNGNPKTKPVITGYR 128
QY 84 --NPNSEESLSKADLIYKDKPEELKKYLKHVHTFVVLNTEGDSDDGLKTIYKDDY 141
DQ 129 AHNVEDVAD--TKGKPLITAKD-----LIKTEFENSND-----YKDLIY 164
QY 142 DKFK-----TPSI---FFV-----FSTKEQEI-----KELFKDK 167
DQ 165 NEFKDYINEELTPSVEEKHFLDDPTLSNGAKGYSPKSDIEIVISDDLSYDMRFTLIHEY 224
QY 168 GNTKERNIAVYNNKNDLHLKFISQYLHQASIFHAVNPYGMPLAATPL-----V 217
DQ 225 AHSQHLNN--DIGKTQISEHRSLSKEIEAESSAYVAVNYGLDTSYSLGVLGSGMGNIS 282
QY 218 DDTVIGKLTAKINFYSLNET--GLDGVPAKGVOLAGGAIDQPTFYHYIKNEAIIEL 275
DQ 283 DDELKAHIK---NIHSPAKTTIEINSLPFSQVID---KKLESEM-----NKDVYKD 329
QY 276 TRINKNNKQN--SKLSALQ-----LSGARDNAYTSATBCLLKRFPVDRGLIYQKNL 325
DQ 330 LSTMTDNLKNGFDKTIYIKGNLVNDYGLNISENSYEND-----DPKNIDYKGF 380
QY 326 SLTISPTQLKLEL-----SVNITNFS-----INAVSLVITQD 360
DQ 381 N-TNNSQDQAKIELINKHDDSLRDYNTFTQTVNRVNNINNTTIFVQDD 427

RESULT 13
ABU26292
ID ABU26292 standard; protein; 598 AA.
XX
AC ABU26292;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #11819.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Campylobacter jejuni.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (BLIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR N-PSDB; ACA30162.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
```

```
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 54216; 1766pp; English.
XX
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 598 AA;

Query Match      6.4%; Score 120.5; DB 6; Length 598;
Best Local Similarity 22.5%; Pred. No. 0.3;
Matches 97; Conservative 69; Mismatches 133; Indels 133; Gaps 25;

QY 10 LLSRIQASRPYNNPLLYVTAKIKYNDKAAKILNLTN---NYEKQIETL----- 60
DQ 171 LLD--LKATFEN-LFKADFPFGAVDNVFKIINNKNISVKNLNNHLENTTHTTTHSMQ 227
QY 61 ---EKDNGNGDQFGKEKTLTKTAMSFPN--SSEESLSKADLIYKDKPEELKKYLKVH 115
DQ 228 KTNKDSG-----SLLSQIKNLDLSLSSKES-----KHENKDKVS--KIE 267
QY 116 RHTFVLINTEGDSDDGLKIYKDDYDKFTPSIFFVFSTKEQ-----EIKELFKD---K 167
DQ 268 EDT-----TDKNTLNKIK--NDEFAKNLTEL--NIKDKKNQDNLNKNESKDLNKFNE 318
QY 168 GNTKERNIAVYNNKD--NLHLKFISQYLHQASIFHAVNPYGMPLAATPLVDDT--VIGK 224
DQ 319 LNKQKKNLNQENTQDQKNLKNNDQNLNLDKNLN-----KEIVKDTQKLVSN 367
QY 225 LRTAKINFYSLNETGLDGVPAKGVOLAGGAIDQPTFYHYIKNEAIIELIRNKK--- 281
DQ 368 LTQKDFN-----LNKEPKNN--NKENDKIKQNFDDQKLNLENKNTQVQV-----NKENN 415
QY 282 ----NNRQNSK-----LSALQSLSGARDNAYTSATBCL 308
DQ 416 ANFNNTNNKETTQEQTKTHSENVNDKNSLDELNSLVKDLNKNYTNQNNARNTPKETLOY 475
QY 309 L---LKRFPVDRGLIYQYK---NLSLTLSPTPQKLELSV-----NITNFSINAVSLVI 356
DQ 476 FSQDLKEAVD---QYKAPITKLISITLNPNNLGEVEVTLIQRGNLHNFNSNANAMNL 530
QY 357 TTQDIVDYQNSL 368
DQ 531 FIQNOAEFKNSL 542
```


CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
SQ Sequence 1331 AA;

Query Match 6.3%; Score 119; DB 6; Length 1331;
Best Local Similarity 21.4%; Pred. No. 1.3;
Matches 85; Conservative 57; Mismatches 140; Indels 116; Gaps 20;

QY 24 NPLIVYKTAJKVNDKAAANYKILNTVNNYKQIETLEKNGNGQDQFGKEKTLTKTAMS 83
Db 788 NPWLSEYL-MPTWQASAS-----NTAFQNFLLKAITINPSQ-----S 825

QY 84 NFFNSSESLKSADLIYIK-DKPEELKKYLKVHRHTFVVLINTEGNSDDGLKIYKDDYD 142
Db 826 QFIQTQDPLTKR--FIYAIDSDKALKN-----NNEQNGSQNHLLT-LNDDFA 869

QY 143 KFKTPSIFVFTSKEQIEKELFKDKGTEKERNIAVYNNKDNHLKFIQYLHQASIFH 202
Db 870 KF-----LYSQF-----ELIKKSGNASNEDLNAIDPENQTIK-DFYNYK-----N 909

QY 203 AVNP--YGMPLAATPLVDDTVIGLRTAKINFYSLNE-----TCLDGVPAFKEGVDLAG 255
Db 910 ALPPLDYKLSENVIGLPKRETIAGQIDTPKYGLTLHGEYQNTPIKIGIKDKWKDKVNLG 969

QY 256 GAIDEQ-----FTYHYIKNEAIIELI-----RIMNKNRQ 285
Db 970 PVLSDNNHIIHQELFKYSPDPLIVNSAAKQYQALGSEINIAVNSFKKIDNKIINQ 1029

QY 286 NSKLGA-LQLSGARDNA-----YTSAECL-----LKRFDVRGLIIQYKNLSLTL 329
Db 1030 DPLVNAFTRVVGINNNSAHDPPEFTSYSTAFKLVLEYPNEWFKKLPFNSF---YANSLLSF 1086

QY 330 SPTPOLKLELSV-NITYNFSINAVSLVITTDIVDYQN 366
Db 1087 VQSTSLFSESGIFPATTSFSTNTVTLVELIKKTIYKN 1124

Search completed: January 24, 2006, 19:44:54
Job time : 85.6852 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:28:54 ; Search time 14.0368 Seconds
(without alignments)
2536.214 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPQDTISVSLDSRIQASRP.....AVSLVITQDIVDYQNSLSA 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	803.5	42.5	356	2	F70212	hypothetical prote
2	134.5	7.1	2295	2	B71621	probable membrane
3	128.5	6.8	1979	2	C71622	hypothetical prote
4	127.5	6.7	1034	2	G90591	hypothetical prote
5	127.5	6.7	1127	2	T28317	ORF MSV156 hypothe
6	126	6.7	913	2	T18503	hypothetical prote
7	126	6.7	1726	2	A45948	major merozoite su
8	126	6.7	2166	2	G70163	hypothetical prote
9	125.5	6.6	1070	2	F90106	IAp100 protein [im
10	121	6.4	1701	2	A54498	major merozoite su
11	121	6.4	1726	1	SAZQGM	major merozoite su
12	121	6.4	2269	2	T28677	rhoetry protein -
13	120.5	6.4	598	2	D81420	hypothetical prote
14	119.5	6.3	834	2	A97178	probable permease
15	119	6.3	612	2	S61116	probable membrane
16	119	6.3	1206	2	A64207	hypothetical prote
17	119	6.3	1411	2	T18417	hypothetical prote
18	118.5	6.3	834	2	T28250	ORF MSV089 probabl
19	117.5	6.2	3394	2	T18501	hypothetical prote
20	117.5	6.2	5005	2	F82884	hypothetical prote
21	117	6.2	1156	2	B70356	chromosome assembl
22	117	6.2	1272	2	H82926	conserved hypotnet
23	117	6.2	1417	2	T18418	hypothetical prote
24	117	6.2	1701	2	A26868	major merozoite su
25	117	6.2	3844	2	T18402	asparagine/aparta
26	116.5	6.2	1191	2	B97116	chromosome segrega
27	116	6.1	1061	2	D98008	conserved hypotnet
28	116	6.1	1367	2	T18466	hypothetical prote
29	115.5	6.1	297	2	T28188	hypothetical prote

ALIGNMENTS

RESULT 1

F70212

hypothetical protein BBA46 - Lyme disease spirochete plasmid A/lp54

C:Species: Borrelia burgdorferi (Lyme disease spirochete)

C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004

C:Accession: F70212

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White

son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,

Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.

Nature 390, 580-586, 1997

A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.

A:Reference number: A70100; MUID:96065943; PMID:9403685

A:Accession: F70212

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-356 <KLE>

A:Cross-references: UNIPROT:O50939; UNIPARC:UPI0000056927; GB:AE000790; NID:g2690224; P1

A:Experimental source: strain B31

C:Genetics:

A:Genome: plasmid

C:Superfamily: Lyme disease spirochete plasmid A/lp54 hypothetical protein BBA46

Query Match 42.5%; Score 803.5; DB 2; Length 356;

Best Local Similarity 46.3%; Pred. No. 4e-45;

Matches 171; Conservative 58; Mismatches 117; Indels 23; Gaps 6;

Qy 1 MPQDTISVSLDSRIQASRPNNYPLVYKTAKIKVKKDAANYKILNLTANNYEKQIETL 60

Db 1 MPKDTISVSFQNRlafnnvnyplliyks-----NVLASHKHALSVTNLQDLLEKL 53

Qy 61 EKDNNGNGDQFGK-----EKLTLKATMSNFFNSSESLKSDLFYIKDPBELKKYKVH 115

Db 54 EMQKAQSDSDSKKIAEQLSALKKSSDFF--GQDGLKSVDFVYV-NQIKDKDFLKN 110

Qy 116 RHTFVLAINTEGNSDGLKIYKDDYDKFTPTSFVFFSTKEQEIKELFKDKNGTEKERN 175

Db 111 LHPFVFNQAE-----VISDFBEIRKICNFIVISTKDANLPDFLKGKDKSELKNI 163

Qy 176 IAVYSNNKDNILHLKPISOYLHQASIFHAVNYPGMLAATPLVDDTVIGKLTAKINIFYSL 235

Db 164 IAVYA-GKQNLHLKFTAAVYHQASIFHAVNYPGMLNAVPYVDDSLDSLAKTNIFYSL 222

Qy 236 LNETGLDVPAPKEGVDLAGGAIDEQFTYHYIKNEAILELIRIWNKNRQSKLSALOLS 295

Db 223 LNETGNDGILAFKGGVSLSGDPIDETFTFYIKNEAIKELIRIWNKNRANSKUGSLND 282

Qy 296 GARDNAYTSAIECLIKRFDRLIIQYKNLSITLSPTQLKLELSVNTYTFNSINAVSIV 355

Db 293 GNLPNYTAGLECHFHKLKQGLIVSYKEIKLINSSEGLSLSEVALKYNDSFNSVNI 342

Qy 356 ITTQDIVDY 364

Db 343 ITTOBINEY 351
||||:|:|
RESULT 2
B71621
probable membrane associated protein PFB0190c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: B71621
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2295 <GAR>
A:Cross-references: UNIPROT:Q9TY98; UNIPARC:UPI000007C03B; GB:AE001379; GB:AE001362; NID
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0190c
Query Match 7.1%; Score 134.5; DB 2; Length 2295;
Best Local Similarity 21.5%; Pred. No. 1.8;
Matches 95; Conservative 57; Mismatches 124; Indels 165; Gaps 23;
QY 23 YNPLLVYKTAIKVNDAAANYKILNLTVNNYEKQIE-TLEKNGNGQDQFGKEKTYLLKTA 81
Db 648 YTLIFNPTKDVFTSQINQIKSKINFENIYEYLITITIKENKNFYDD-----LLKCR 701
QY 82 MSNFFNSSSELSKADLFYKDKPELKK-----YLVKVRHTFVVLINTEGNSDDGLKIY 137
Db 702 REIFF-KRHLNENIM--DQELKKNIRNLRIHE-----VSNEGNRNNTINRKY 751
QY 138 KD-----DYDKFTPSIFFVFSTKEQIEKEL-----FKDKGNTKERN-----IAVY 179
Db 752 KKGTVDYDKM--ELYVV-----EKNILNVNDTNTFNFNKKEKDKNYFDINKTWRIYDY 805
QY 180 SNN-----KQNLH--LK-----FTSQVLHQA-----SIFHAVN-----205
Db 806 YNNINLNIPTPAIRKMKDKIYDQLKLLRSNFVEKLNKESICVLSFLYLIIGINDNGKLHF 865
QY 206 PYGMPAAATPLVDVTVIGKRTAKNFYSLNLTGLDGVPAFKEG-----251
Db 866 PYGFP-----RNIDFSVKLIREGDKGLCNFLSGVLVHINLPIFVNSS 908
QY 252 -----DLAGGAIDEQFTYHYIKNEAI-----IELIRIWNKNRQNSKLSAL 292
Db 909 ISISTEMNDVLEMDNSINSFPIYIYKKNENIRNHDFLSDENRIIPRKEDNIKSKIISY 968
QY 293 QLSGARDNAYTSAECLLKRFDVRGLIIOYKNLSUTLSTPQKLELSVNITY-----NF 347
Db 969 SLGSKDDPFSL-----AFTNNVIRLKYKN-----KTNNTYLDKDYDF 1007
QY 348 SINAVSLVTTQDIDVDYQNSL 368
Db 1008 -----TFDKINYNKSV 1018
RESULT 3
C71622
hypothetical protein PFB0145c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: C71622

A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1979 <GAR>
A:Cross-references: UNIPROT:O96133; UNIPARC:UPI0000007EA68; GB:AE001375; GB:AE001362; NID
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0145c
Query Match 6.8%; Score 128.5; DB 2; Length 1979;
Best Local Similarity 20.9%; Pred. No. 3.6;
Matches 77; Conservative 57; Mismatches 128; Indels 107; Gaps 16;
QY 3 QDTISVSLDSRIQASR-----PNYINPLL-----VYKTAK--IKVNDKDAAN-- 42
Db 242 RETISIEIIDIKKHLEKIKIEKKEDLENLKKLLSKENVLKELKGVCKEKNETINSL 301
QY 43 -----YKILNLTVNNYEKOIETLEKNGNGQDQFGKEKTYLLKTAMSNFFNSSSES 92
Db 302 NDNIIEKEKKYKLLLEYELEENKQIDLLNKOKEKEKEKEKEK-----EREKEK 352
QY 93 LKSADLFYKDKPELKKYLKVH-----RHTFVVLINTEG-----NSDDGL 134
Db 353 EKEYDTLIKELDKESISILEKHSIKVREMDIEKREHNFLEHWDQLKDKNSFVKNQOL 412
QY 135 KIYKDDYDKFTPSIFFVFSTKEQIEKEL-----FKDKGN-----TEKERNIAVYS--NNKD 184
Db 413 KVKYCEIKNLKTE-----LEKKEKELDIENVSKKEINKLINQLNEKEKEKQILAFKNHKE 467
QY 185 NLH-LKFTSQVLHQA-SIFHAVNPGMPLAATPLVDVTVIGKRTAKNFYSLNLTGLDQ 243
Db 468 BIHGLK-----BELKESVKITKITQEOLEQVMDIKQKELDQ 503
QY 244 VPAKEGVDLAGGAI-----EQFTYHYIKNEAI--IELIRIWNKNRQNSKLSALQLSGARDN 300
Db 504 LQE-KYNAQIESISIELSKKEKEYNQYKNTYIEINLNLEKLEETNKEYTTLNQL-----N 556
QY 301 AYTSIAECL 309
Db 557 NYTWEINML 565
RESULT 4
G90591
hypothetical protein MYPU 6390 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: G90591
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A99512; MUID:21267165; PMID:111353084
A:Accession: G90591
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1034 <KUR>
A:Cross-references: UNIPROT:Q98PT0; UNIPARC:UPI000000D4633; GB:AL445566; PID:gl4090054; P
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYPU 6390
A:Genetic code: SGC3
Query Match 6.7%; Score 127.5; DB 2; Length 1034;
Best Local Similarity 23.8%; Pred. No. 1.8;
Matches 94; Conservative 67; Mismatches 141; Indels 93; Gaps 21;
QY 32 AKIKVNDKDAANYKIL-----NLTVNNYEKOIETLEK-NGNGQDQFGKEKTYLLKTAMS--- 83
Db 227 AQVKISKKEIKNKVLEDKNNLKI-KYKDRINKAEYSYNDYFEQNKQIKESLKHANEXIA 285
QY 84 -----NFFNSSE-ESLKS-----ADLFYKDKPEELKKYLKVH-RHTFVVLINT 125
Db 286 QNKDKINELNKNKLEKIKALAEKRALKNYLYFKKLIILAKKYKLYLYINHOKVLLICN 345

QY 126 --EGDNS--DGLKLYKDDYDKFTPTSPFFVFTSTKEQBIK-----ELPK-----DKGNT 171
 Db 346 YYEKDKTQIDQLWKYKXN-----FLNKVDDQVQNDYFELYKKIINVYENDNSY 395
 QY 172 KERNIAVYNNKDNHLKPFISOYLHOASIFRAVNPYGMPLAATPLVDVTVIGKLTAKIN 231
 Db 396 KKQVVKLIINKSYNPLIKY---ELKKNALFHLKSHQWQNL-----LIKQSSYEGD 444
 QY 232 FYSL-----LNETGLDGVPAFKEGVDLAGAIDQFTYHYIKNEAIIELIRIWNKNRQ-- 285
 Db 445 FVEVKSALSOYVIDSYNEINNSIIEKQKILEELFNQYAKN-----NLENKQVF 494
 QY 286 NKSLSALQISGARDNAYTSATECLKR-----FVDRGLIIOYKML--SLTI-SPTP 333
 Db 495 QSKVNLKVDYLLEKA---KVKLAKKEITKALVFTNKKLKIDFKEKVNLSKLENPKY 551
 QY 334 QLKLELSVNIITVNFISINAVSLVITTDIVDVQNSL 368
 Db 552 QNKLLKTNLGRFLSKKLLHKIYOSKILEAQSI 586

RESULT 5
 T28317
 ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
 C:Species: Melanoplus sanguinipes entomopoxvirus
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
 C:Accession: T28317
 R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
 J. Virol. 73, 533-552, 1999
 A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
 A:Reference number: 220484; MUID:99102612; PMID:9847359
 A:Accession: T28317
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1127 <AFO>
 A:Cross-references: UNIPROT:Q9YVT6; UNIPARC:UPI00000F6900; EMBL:AF063866; NID:g4049647;
 C:Genetics:
 A:Note: MSV156

Query Match 6.7%; Score 127.5; DB 2; Length 1127;
 Best Local Similarity 21.1%; Pred. No. 2;
 Matches 96; Conservative 68; Mismatches 142; Indels 149; Gaps 21;

QY 32 AKIKVNDAAANYKILNLT---VNNYEKQIETLEKNGQDQFGKEKTLTKTAMSNFFNS 88
 Db 662 ASITNNIENISNKIKDLNEFIISNDSSEKLEDEIRKYQ-QFDKIKAMNTEVKSFNT 720
 QY 89 SEESLSAD-----LFYDKPEELKYLKVHRTFVVLINTEGD-----NSDD 132
 Db 721 LQKDIDSISKNINELTNAYDIINTKANDLDDKLNNGYSEFKNLYNNSADLDDTIQKNDE 780
 QY 133 GLKIKYDDYDKPTPSI-----FFVFTKE-----QETKELFK 165
 Db 781 KVKQLNEVLEKKNQKQSIENDIVNNFIKELIKFNNTNKSLSNELLTNDINDIKIFLYK 840
 QY 166 DKGNTKERNIAVYNNKDNL-----HLKPFISOYL-----HQASIFHAVNPYGMPLA 212
 Db 841 ELNKISTNNLLIKYKNEIDNVNNEKLSIVENIQFINSFLSIEFNOGSITSINFLNTLA 900
 QY 213 APPLVDDTVIGKL-----RTAKINFYSLNLTGDLGVPA 246
 Db 901 G---IND-VINKNLKIMADTTRGDTNIRDBIKNQISSENKSKQFNKNEKDKLKLIS 956
 QY 247 FKEGV---DLAGG-----AIDQFTYHYIKNEAIIELIRIWNKNRQN 286
 Db 957 FNDKLNKYNISAGYTEYNNIEHECLKYLYAVSDQ-EYRYK--FLIHNKIQDK-NKN 1012
 QY 287 SKLSALQISGARDNAYTSIECLLRFDVDRGLIIOYKNSLTLSPPLK----- 336
 Db 1013 SMLQFLKL-----FLKRYI---IYDLKILSDTDI-NIQIKOYLYNKLKN 1054
 QY 337 -----LELSVNITNFSINAVSLV-IITQDIVDY 364

Db 1055 EYKLTPEQFINQTYGFTMPIFNILINISTNMVFDY 1089

RESULT 6
 T18503
 hypothetical protein C0750w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T18503
 R:Lawson, D.; Bowman, S.; Barrell, B.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18503
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-913 <LAW>
 A:Cross-references: UNIPROT:O77386; UNIPARC:UPI000017CC3C; EMBL:Z98551; NID:e1331903; PI
 C:Genetics:
 A:Map position: 3
 A:Introns: 384/3
 A:Note: C0750w

Query Match 6.7%; Score 126; DB 2; Length 913;
 Best Local Similarity 23.1%; Pred. No. 1.9; 130; Indels 64; Gaps 15;
 Matches 74; Conservative 52; Mismatches 52; Indels 64; Gaps 15;

QY 8 VSLDSRIQASRPVNPYLLVYKTAK-----IKVNDAAANYKILNLTVNNYEKQIETLEK 62
 Db 510 VSLYD---VLSRSNAHG--LIYIDGKENNTSINLNNKNNNNNNNNNNNNNNIS 564
 QY 63 DNGNGQDQFGKEKTLTKTAMSNFFNSSELSKADLFYKDKPBLK-KYLKVHRTFV 121
 Db 565 NNSNNSNFGHEKVI-TIIPYGPISPTIKIE--FLNKQNNNEKWAYLRLOKNEY-I 620
 QY 122 LINTEGNSDP-----GLKIYKDDYDKPK-----TPSIFVFTSTKEQIETLEK 169
 Db 621 MKKVGINSSDIFPLSYKYKYKQEKIFILNDFSETTYLQFICQTKSEIKSLCTISLV 680
 QY 170 TEKERNIAVYNNKDNHLKPFISOYLHOASIFHAVNPYGMPLAATPLVDVTVIGKLTAK 229
 Db 681 DSHKNNMAINLFTNNH---ILFLYDQSPRKIQPY-----NQIISFRFVL 726
 QY 230 I-NFYSLNLTGDLGVPA-----FKEGVDLAGAIDQFTYHYIKNEAIIELI 276
 Db 727 FKNFHLFQENKINVILLKCMCSSSSSYNTQKNIELSYLTNEQVSHDFIHTPGI--LI 784
 QY 277 R-----IWNKNRQNSK 288
 Db 785 RPRTNISHNDYQKNKGNHK 804

RESULT 7
 A45948
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Dec-2004
 C:Accession: A45948
 R:Chang, S.P.; Kramer, K.J.; Yamaga, K.M.; Kato, A.; Case, S.E.; Siddiqui, W.A.
 Exp. Parasitol. 67, 1-11, 1998
 A:Title: Plasmodium falciparum: gene structure and hydropathy profile of the major merozo
 A:Reference number: A45948; MUID:89005525; PMID:3049134
 A:Accession: A45948
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1726 <CHA>
 A:Cross-references: UNIPROT:Q25922; UNIPARC:UPI0000177F89; GB:M37213
 C:Superfamily: G surface protein
 C:Keywords: surface antigen

Query Match 6.7%; Score 126; DB 2; Length 1726;
 Best Local Similarity 21.9%; Pred. No. 4.3;
 Matches 94; Conservative 64; Mismatches 145; Indels 126; Gaps 22;

A54498
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (is
 C:Species: Plasmodium falciparum
 C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 31-Dec-2004
 C:Accession: A54498
 R:Peterson, M.G.; Coppel, R.L.; McIntyre, P.; Langford, C.J.; Woodrow, G.; Brown, G.V.;
 Mol. Biochem. Parasitol. 27, 291-302, 1988
 A:Title: Variation in the precursor to the major merozoite surface antigens of Plasmodium
 A:Reference number: A54498; MUID:88142999; PMID:2449612
 A:Accession: A54498
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1701 <P>
 A:Cross-references: UNIPROT:P13819; UNIPARC:UPI000012F631; GB:M19143; NID:gl60412; PIDN:
 C:Superfamily: G surface protein
 C:Keywords: surface antigen

Query Match 6.4%; Score 121; DB 2; Length 1701;
 Best Local Similarity 21.7%; Pred. No. 9;
 Matches 93; Conservative 64; Mismatches 146; Indels 126; Gaps 22;

QY 5 TISVSLD---SRIQASRPNYN---PLLVTAKI-----KVNKDAANYKILNL----- 48
 DB 974 TISTTEMEKFYENILKNNDTYFNDDIKQFVKSNSKVITGLTETQKNALNDEIKKLKDTLQ 1033
 QY 49 ----TANNYKQIETL-EKONGGQDQFG-KEKTLTKTAMS---NFFNSSEESLSADLF 99
 DB 1034 LSFDLVNYKYLKDLRFLNKKELGQDKMQIKLTLLEQLESKLSLNNPHNVLFQNSVF 1093
 QY 100 IYKDPBELKYLKVRHRTFVVLINTEG-----DNSDDGLKIY-----KDDY---DKF 144
 DB 1094 FNKKEAEIAETENTLENTKILLKHYGLVYNGESSPLKTLSEVSIQTEDNYANLEKF 1153
 QY 145 KTPS-----IPFVSTKEQ---EIKELFKDK---GNTEKERNIAVYNN 182
 DB 1154 RALSKIDGKLNLDNLHLGKKLSFLSSGLHLHILTELKEVKNKNTYGNPSPE-----NNK 1207
 QY 183 KDNHLKFTSOYLHOASIFHAVNPYGMPLAATPL-----VDDTVI---GKLRTA 228
 DB 1208 KNEALKSVENFLPEAKVTVTTPQDPVTPSPVSRVSGSGSTKEETQIPTSGSLITE 1267
 QY 229 KINFVSLN-----ETGLDGPAPKGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNR 284
 DB 1268 LQQVVLQNYDEEDSLVLPFGESED-----NDEYLDQVVTGEAI----- 1309
 QY 285 QNSKLSALQSGARDNAYTSAIECLLKRFDVDRGLIIQYKNLSLTLSPTPQKLKLSVNI-I 343
 DB 1310 -----SVTMDN-----ILSGPENEDYVILKPLA---GVYRSLKKQIEKNII 1348
 QY 344 TYNFSINAV 352
 DB 1349 TPNLNLNDI 1357

RESULT 11
 SAZQGM
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
 N:Alternate names: 195K glycoprotein
 C:Species: Plasmodium falciparum
 C:Date: 30-Sep-1987 #sequence_revision 31-Mar-1991 #text_change 31-Dec-2004
 C:Accession: A23386; S06361
 R:Weber, J.L.; Leininger, W.M.; Lyon, J.A.
 Nucleic Acids Res. 14, 3311-3323, 1986
 A:Title: Variation in the gene encoding a major merozoite surface antigen of the human m
 A:Reference number: A23386; MUID:86205236; PMID:3517809
 A:Accession: A23386
 A:Molecule type: DNA
 A:Residues: 1-1104 <WEB1>
 A:Cross-references: UNIPROT:P04934; UNIPARC:UPI0000174696; EMBL:X03831
 R:Weber, J.L.; Sim, B.K.L.; Lyon, J.A.; Woliff, R.
 Nucleic Acids Res. 16, 1206, 1988
 A:Title: Merozoite surface protein sequence from the Camp strain of the human malaria pa
 A:Reference number: S06361; MUID:88143999; PMID:3278296

A:Accession: S06361
 A:Molecule type: DNA
 A:Residues: 1104-1726 <WEB2>
 A:Cross-references: UNIPARC:UPI0000174697; EMBL:X03831
 C:Comment: The merozoite stages of different strains have strain-specific surface antigen
 C:Superfamily: G surface protein
 C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat
 F:1-19/Domains: signal sequence #status predicted <SIG>
 F:20-1726/Product: major merozoite surface antigen #status predicted <MAT>
 F:67-87,91-96,100-105,109-120/Region: 3-residue repeats (S-G-T)
 F:757-765/Region: 3-residue repeats (T-E-E)
 F:133,272,501,567,638,827,839,924,944,990,1016,1114,1221,1613,1658/Binding site: carboxy

Query Match 6.4%; Score 121; DB 1; Length 1726;
 Best Local Similarity 21.7%; Pred. No. 9.2;
 Matches 93; Conservative 64; Mismatches 146; Indels 126; Gaps 22;

QY 5 TISVSLD---SRIQASRPNYN---PLLVTAKI-----KVNKDAANYKILNL----- 48
 DB 999 TISTTEMEKFYENILKNNDTYFNDDIKQFVKSNSKVITGLTETQKNALNDEIKKLKDTLQ 1058
 QY 49 ----TANNYKQIETL-EKONGGQDQFG-KEKTLTKTAMS---NFFNSSEESLSADLF 99
 DB 1059 LSFDLVNYKYLKDLRFLNKKELGQDKMQIKLTLLEQLESKLSLNNPHNVLFQNSVF 1118
 QY 100 IYKDPBELKYLKVRHRTFVVLINTEG-----DNSDDGLKIY-----KDDY---DKF 144
 DB 1119 FNKKEAEIAETENTLENTKILLKHYGLVYNGESSPLKTLSEVSIQTEDNYANLEKF 1178
 QY 145 KTPS-----IPFVSTKEQ---EIKELFKDK---GNTEKERNIAVYNN 182
 DB 1179 RVLKIDGKLNLDNLHLGKKLSFLSSGLHLHILTELKEVKNKNTYGNPSPE-----NNK 1232
 QY 183 KDNHLKFTSOYLHOASIFHAVNPYGMPLAATPL-----VDDTVI---GKLRTA 228
 DB 1233 KNEALKSVENFLPEAKVTVTTPQDPVTPSPVSRVSGSGSTKEETQIPTSGSLITE 1292
 QY 229 KINFVSLN-----ETGLDGPAPKGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNR 284
 DB 1293 LQQVVLQNYDEEDSLVLPFGESED-----NDEYLDQVVTGEAI----- 1334
 QY 285 QNSKLSALQSGARDNAYTSAIECLLKRFDVDRGLIIQYKNLSLTLSPTPQKLKLSVNI- 343
 DB 1335 -----SVTMDN-----ILSGPENEDYVILKPLA---GVYRSLKKQIEKNIF 1373
 QY 344 TYNFSINAV 352
 DB 1374 TPNLNLNDI 1382

RESULT 12
 T28677
 rhoptry protein - Plasmodium yoelii
 C:Species: Plasmodium yoelii
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T28677; C45521
 R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
 Mol. Biochem. Parasitol. 65, 171-177, 1994
 A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.
 A:Reference number: Z20508; MUID:95021522; PMID:7935623
 A:Accession: T28677
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2269 <KEE>
 A:Cross-references: UNIPROT:Q26223; UNIPARC:UPI0000080871; EMBL:L27838; NID:g457145; PID
 R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cc
 A:Reference number: A45521; MUID:91101660; PMID:2270106
 A:Accession: C45521
 A:Status: preliminary
 A:Molecule type: DNA

A;Residues: 2131-2269 <KE2>
A;Cross-references: UNIPARC:UPI000017B646; GB:M34283

Query Match 6.4%; Score 121; DB 2; Length 2269;
Best Local Similarity 19.8%; Pred. No. 13;
Matches 78; Conservative 69; Mismatches 138; Indels 108; Gaps 17;

QY 29 YKTIKIVKVK-----DAANYKILNLTNVNVE-----KQETLE-----K 62
DB 1114 FTKLIKINEIPKSDCLKETNDLEKQISNLSIDTQETKLTENGKQKLTLELLESKK 1173

QY 63 DNGNGQDQFGKEKTLTKTAMSNFFNSSELSKADLFY-----KQPEELKKYL 112
DB 1174 OKKNIEDQ-KKELDEVNSKIKNIENVQHKKNGYEGIVEKINEIAKTNNQIESTKELI 1232

QY 113 KVRHTFVVLINT---EGNSDDGLKIYKOD-----YDKF-KTPSIF--FVSTKEQEI-- 160
DB 1233 KPTIQHIISSFNANDEGIDSDENLGKYNTEMGNTYEEFIKSYNLITNLYLETVSKESITY 1292

QY 161 KFLKDKGNTEKERNIAVYNNKDNHLKLFISQYLHQASIFHAVNPYGMPLAATPLVDVT 220
DB 1293 NOIQNKRIDTQKELLKNTEVNVKAKSYLDYIKENEFDRIVTHFKK----- 1337

QY 221 VIGKLTAKINF---YSLNETGLDGPAPKEGVDLAGGAIDEQ----- 261
DB 1338 ---KLNTVNDNPKNEYSKVE-GFDNI---SNSINTVKNSTDSLLNLNQTKEMYANI 1390

QY 262 -----FTHYIKVEAILIRIWNKNNQNSKLSALQLSGARDNAYTSAIECLLKRFVDR 316
DB 1391 VNNTYSKYAEAFNIPKPLANTLNITKINSSGIDLS-----KDIKAILSYLD- 1441

QY 317 GLIIQVKNLSLSTPQKLSLVNITVNSFI 349
DB 1442 ----SKTEDLFIPIFPQKTTYTKISDSYSI 1470

RESULT 13
D81420
hypothetical protein Cj0041 [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: D81420
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: D81420
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-598 <PAR>
A;Cross-references: UNIPROT:Q9PU70; UNIPARC:UPI00000C2013; GB:AL139074; GB:AL111168; NID
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0041

Query Match 6.4%; Score 120.5; DB 2; Length 598;
Best Local Similarity 22.5%; Pred. No. 2.4;
Matches 97; Conservative 69; Mismatches 133; Indels 133; Gaps 25;

QY 10 LLDSDRIQASRPNNYPNLLVYTKAKIVNKDAAANYKILNLTVN---NYEKQIETL----- 60
DB 171 LLD--LKATFPN-LDKADFPKGVADVNFKEIINNKSIVSKNKNLENLTHTTTHSMQ 227

QY 61 ---EKDNGNGQDQFGKEKTLTKTAMSNFFN---SSELSKADLFYKQPEELKKYLKVH 115
DB 228 KTNKDSG-----SLLSQTLKNLDSILSSKES-----KHEKNDKVK-S-XIE 267

QY 116 RHTFVVLINTEGNSDDGLKIYKDDYDKETPSIFVSTKEQ-----EIKELFKD--K 167
DB 268 EDT-----TDAKNTLNKIK--NDEFKVLNTEEL-NIKKQKQDNLNKESKDLNKFKNKE 318

QY 168 GNTKERNIAVYNNKDNHLKLFISQYLHQASIFHAVNPYGMPLAATPLVDVT--VIGK 224

Db 319 LNKQEKNNLNQENIQDNKNLKNNDQNLNLDKNLN-----KEIVKDTQKLVSN 367
QY 225 LRTAKINPYSLLNETGLDGPAPKEGVDLAGGAIDEQFTYHYIKVEAILIRIWNK--- 281
DB 368 LTQKDFN---LNKEPKNN---NKENKDIKQNFDDQKLNFNENLNKTQVVQ-----NKENN 415

QY 282 ---NNRNSK-----LSALQLSGARDNAYTSAIEC 308
DB 416 ANFNNTNNKETTQEGTKTHSENVKNSLDELNSLVKDLNKNVTONNARITPKETIQY 475

QY 309 L---LKRFPVDRGLIIQYK---NLSLTSPPTPQKLELSV-----NITYNFSINAVSLVI 356
DB 476 FSQDLKEAVD---QYKAPITKLSITLNPNNLGEVETLTIQRGNNLHNFNSNANAMNL 530

QY 357 TTQIDIVYQNSL 368
DB 531 FIONQAEPKNSL 542

RESULT 14
A97178
probable permease [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: A97178
R;Nolling, J.; Breston, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97178
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-834 <KUR>
A;Cross-references: UNIPROT:Q97GW0; UNIPARC:UPI00000CA47D; GB:AE001437; PIDN:AAK80212.1;
A;Experimental source: Clostridium acetobutylicum ATCC924
C;Genetics:
A;Gene: CAC2255

Query Match 6.3%; Score 119.5; DB 2; Length 834;
Best Local Similarity 23.5%; Pred. No. 4.4;
Matches 92; Conservative 62; Mismatches 137; Indels 101; Gaps 22;

QY 6 ISVSLDSDRIQASRPNNYPNLL--VYKTAKIVNKD--AANYKIL---NLTVNNEYKQIE 58
DB 480 ISISLMTIMYMTNSLSKDLTLEYIKNLNFDVSDIDINSQNVVDVLKCKSEIINNYYDKTN 539

QY 59 TLEKD--NGNGQDQFGKEKTLTKTAMSNFFNSSELSKADLF-IYKDKPEELKKYLKVH 115
DB 540 VIRKYYINTNLNDDFSKKIDLLCVDPNNF-----KSYDNYLVYTNKNKQLDQL--- 587

QY 116 RHTFVVLINTEGNSDDGLKIYKOD---YDKETPSIFVSTKEQ-----IKELFK 165
DB 588 ---NNNEDGIIISSEKNSRRYNNLKGDTISLYSSNSEKEDFKILSVVNAGFM 635

QY 166 DKGNT-----KERNIAVYNNKDNHLKLFISQYLHQASI-FHAVNPY-GMPLAATP 215
DB 636 NMGNINLISFKAALKHFNR-YSNE-----FFISTSTNTSSIKKDLKKFKGFPVSVST 688

QY 216 LVDDTVIGKLTAKINFYSLNETGLDGPAPKEGVDLAGGAIDEQFTYHYIKVEAILIR 275
DB 689 KQDN--VNKDAASSKEFISLFN-----IF-----SFTSMIALIGILSN 725

QY 276 IRI--WNKNNQNSKLSALQLSGARDNAYTSAIECLLKFPVDRGLIIQYKNL-----SLTL 329
DB 736 ISISFVQRKEIASLISIGLSSTGKN-----LMILFESFTE-GLIAFTFLSSITWTLTL 779

QY 330 SPTQKLELSVNITYNFSIN-----AVSLVI 356
DB 780 LSDIFKYLVLNITNITYPFSPFWAALSVSIIIM 811

Search completed: January 24, 2006, 19:59:16
Job time : 15.0368 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:27:53 ; Search time 79.505 Seconds
(without alignments)
3283.383 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPQDTISVSLDSRIQASRP.....AVSLVITQDIVYQNSLSA 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1889	100.0	370	Q9S096 BORBU	Q9S096 borrelia bu
2	1884	99.7	370	Q5XYW3 BORG	Q5XYW3 borrelia bu
3	1878	99.4	370	Q9S0C8 BORBU	Q9S0C8 borrelia bu
4	1837	97.2	370	Q9R2W7 BORBU	Q9R2W7 borrelia bu
5	1829	96.8	370	Q9S003 BORBU	Q9S003 borrelia bu
6	1828	96.8	370	Q9S0F9 BORBU	Q9S0F9 borrelia bu
7	1755	92.9	362	Q9S0S1 BORBU	Q9S0S1 borrelia bu
8	1582.5	83.8	373	Q5XZ68 BORBU	Q5XZ68 borrelia ga
9	803.5	42.5	356	O50939 BORBU	O50939 borrelia bu
10	421	22.3	106	O5XYV3 BORG	O5XYV3 borrelia ga
11	382.5	20.2	156	Q6AS11 BORG	Q6AS11 borrelia ga
12	140.5	7.4	76	Q6AS12 BORG	Q6AS12 borrelia ga
13	140	7.4	2670	Q7YZ02 CRYP	Q7YZ02 cryptospori
14	140	7.4	2711	O5CWH1 CRYP	O5CWH1 cryptospori
15	137	7.3	1057	O7RB65 PLAYO	O7RB65 plasmodium
16	136	7.2	1158	Q7RC65 PLAYO	Q7RC65 plasmodium
17	136	7.2	1356	Q5ZZZ2 MYCH	Q5ZZZ2 mycoplasma
18	135.5	7.2	654	Q9RLP9 MYCH	Q9RLP9 mycoplasma
19	134.5	7.1	1347	Q4Z0S9 PLABE	Q4Z0S9 plasmodium
20	134.5	7.1	2295	Q9TY98 PLAF7	Q9TY98 plasmodium
21	134	7.1	872	Q5TIA9 HUMAN	Q5TIA9 homo sapien
22	134	7.1	875	Q8NDG7 HUMAN	Q8NDG7 homo sapien
23	134	7.1	1006	Q4YZ95 PLABE	Q4YZ95 plasmodium
24	134	7.1	1012	Q5TIB0 HUMAN	Q5TIB0 homo sapien
25	134	7.1	1309	O8TJEB PLAF7	O8TJEB plasmodium
26	133.5	7.1	1093	Q5ANT5 DICI	Q5ANT5 dictyosteli
27	133.5	7.1	1262	Q4YSN7 PLABE	Q4YSN7 plasmodium
28	132.5	7.0	1464	Q8TIF6 PLAF7	Q8TIF6 plasmodium
29	132.5	7.0	1633	Q8CMP4 STABP	Q8CMP4 staphylococ
30	131.5	7.0	984	Q86IF8 DICI	Q86IF8 dictyosteli
31	131.5	7.0	996	Q559A5 DICI	Q559A5 dictyosteli

32 131.5 7.0 1186 2 Q895M7 CLOTE Q895M7 clostridium

33 131 6.9 867 1 IF2 BUCAP Q8K9H1 buchnera ap

34 130.5 6.9 798 2 Q7RQ74 PLAYO Q7RQ74 plasmodium

35 130.5 6.9 1161 2 Q54NR5 DICI Q54NR5 dictyosteli

36 130.5 6.9 1733 2 Q9K114 STABP Q9K114 staphylococ

37 130.5 6.9 2226 2 Q97225 PLAF7 Q97225 plasmodium

38 129.5 6.9 943 2 Q8TLY0 PLAF7 Q8TLY0 plasmodium

39 129.5 6.9 1877 2 Q9XXW1 9APIC Q9XXW1 plasmodium

40 129.5 6.9 1879 2 Q7PDP1 PLAYO Q7PDP1 plasmodium

41 129.5 6.9 1936 2 Q8IDT8 PLAF7 Q8IDT8 plasmodium

42 129 6.8 456 2 Q4YRA3 PLABE Q4YRA3 plasmodium

43 129 6.8 565 2 Q4U973 THEAN Q4U973 theileria a

44 129 6.8 1046 2 Q4YRU4 PLABE Q4YRU4 plasmodium

45 128.5 6.8 373 2 Q630X2 BACCZ Q630X2 bacillus ce

ALIGNMENTS

RESULT 1

ID Q9S096 BORBU PRELIMINARY; PRT; 370 AA.

AC Q9S096;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein BBO11.

GN OrderedLocustNames=BBO11;

OS Borrelia burgdorferi (Lyme disease spirochete).

OG Plasmid cp32-7.

OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;

OC Borrelia burgdorferi group.

OK NCBI_TaxID=139;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ATCC 35210 / B31;

EX MEDLINE=20138354; PubMed=10672174;

RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,

RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,

RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;

RT "A bacterial genome in flux: the twelve linear and nine circular

RT extrachromosomal DNAs in an infectious isolate of the Lyme disease

RT spirochete Borrelia burgdorferi.";

RL Mol. Microbiol. 35:490-516(2000).

[2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ATCC 35210 / B31;

EX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,

RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,

RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,

RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,

RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,

RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,

RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochaete, Borrelia

RT burgdorferi.";

RL Nature 390:580-586(1997).

DR EMBL; AE001579; AAF07607.1; -; Genomic_DNA.

DR TIGR; BBO11; -.

DR InterPro; IPR008505; DUF787.

DR Pfam; PF05619; DUF787; 1.

KW Complete proteome; Hypothetical protein; Plasmid.

SQ SEQUENCE 370 AA; 42103 MW; 62FB9015CFEBCD6 CRC64;

Query Match 100.0%; Score 1889; DB 2; Length 370;

Best Local Similarity 100.0%; Pred. No. 4.2e-111;

Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MPQDTISVSLDSRIQASRPYNPLLYVTAKIKVKKDAANYKILNLTNNYKQIETL 60

Db 1 MPQDTISVSLDSRIQASRPYNPLLYVTAKIKVKKDAANYKILNLTNNYKQIETL 60

Db 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKASADLFYKDKPEELKKYLVKVRHTFV 120
Qy 121 VLINTEGNSDDGLKIYKDDYKFTPSIFFVSTKEQEIKEKELFKDKGNTKERNIAVYS 180
Db 121 VLINTEGNSDDGLKIYKDDYKFTPSIFFVSTKEQEIKEKELFKDKGNTKERNIAVYS 180
Qy 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
Db 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
Qy 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
Db 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
Qy 301 AYTSIAIECLLKRFDVDRGLIIOYKNLSLTSPQPKLELSVNTYFNFSINAVSLVITTD 360
Db 301 AYTSIAIECLLKRFDVDRGLIIOYKNLSLTSPQPKLELSVNTYFNFSINAVSLVITTD 360
Qy 361 IVDYQNSLSA 370
Db 361 IVDYQNSLSA 370

RESULT 4

Q9R2W7 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9R2W7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein BBL11 (Hypothetical protein BBL11).
GN OrderedLocNames=BBL11, BBL11, BBS11;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-8, Plasmid cp32-1, and Plasmid cp32-3.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=CP32-8, CP32-1, and CP32-3;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31; PLASMID=CP32-8, CP32-1, and CP32-3;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AE001575; AAF07434.1; -; Genomic DNA.
DR EMBL; AE001576; AAF07479.1; -; Genomic DNA.
DR EMBL; AE001580; AAF07651.1; -; Genomic DNA.
DR TIGR; BBL11; -;
DR TIGR; BBL11; -;
DR TIGR; BBS11; -;
DR TIGR; BBS11; -;
DR IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.

KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 370 AA; 42102 MW; D071E1723FA9767C CRC64;
Query Match 97.2%; Score 1837; DB 2; Length 370;
Best Local Similarity 97.0%; Pred. No. 8e-108;
Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MPQDTISVSLDSSRIQASRPNNYNPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
Db 1 MPQDTISVSLDSSRIQASRPNNYNPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL 60
Qy 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKASADLFYKDKPEELKKYLVKVRHTFV 120
Db 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKASADLFYKDKPEELKKYLVKVRHTFV 120
Qy 121 VLINTEGNSDDGLKIYKDDYKFTPSIFFVSTKEQEIKEKELFKDKGNTKERNIAVYS 180
Db 121 VLINTEGNSDDGLKIYKDDYKFTPSIFFVSTKEQEIKEKELFKDKGNTKERNIAVYS 180
Qy 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
Db 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
Qy 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
Db 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
Qy 301 AYTSIAIECLLKRFDVDRGLIIOYKNLSLTSPQPKLELSVNTYFNFSINAVSLVITTD 360
Db 301 AYTSIAIECLLKRFDVDRGLIIOYKNLSLTSPQPKLELSVNTYFNFSINAVSLVITTD 360
Qy 361 IVDYQNSLSA 370
Db 361 IVDYQNSLSA 370
RESULT 5
Q9S003 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9S003;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein BBO18.
GN OrderedLocNames=BBO18;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp56.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RT spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;

RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";
RL Nature 390:580-586 (1997).
DR EMBL: AE001584; AAF07741.1; -; Genomic_DNA.
DR TIGR: BBQ18; -;
DR InterPro: IPR008505; DUF787.
DR Pfam: PF05619; DUF787; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 370 AA; 42150 MW; DABE039E8F8EBBD8 CRC64;

Query Match 96.8%; Score 1829; DB 2; Length 370;
Best Local Similarity 96.8%; Pred. No. 2.6e-107;
Matches 358; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 MPQDTISVSLDSRIQASRPNNYPLLYVYKTAKIKVNDAAANYKILNLTVNNYKQIETL 60
DB 1 MPQDTISVSLDSRIQASRPNNYPLLYVYKTAKIKVNDAAANYKILNLTVNNYKQIETL 60
QY 61 EKXNGNGQDFGKEKTLTKTAMSNFFNSSESLKSADLFYKDKPEELKYLKVHRHTFV 120
DB 61 EKXNGNGQDFGKEKTLTKTAMSNFFNSSESLKSADLFYKDKPEELKYLKVHRHTFV 120
QY 121 VLINTEGNSDGLKIYKDDYDKPTSPFFVSTKEQEIKELPKDKGNTKERNIAVYS 180
DB 121 VLINTEGNSDGLKIYKDDYDKPTSPFFVSTKEQEIKELPKDKGNTKERNIAVYS 180
QY 181 NNKDNHLKFIQYHLQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
DB 181 NNKDNHLKFIQYHLQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
QY 241 LDGVPAPFEGVDLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQSGARDN 300
DB 241 LDGVPAPFEGVDLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQSGARDN 300
QY 301 AYTSAEICLLKRFVDRGLIIQYKNSLTSPTPQLKLSVNTYFNFSINAVSLVITQD 360
DB 301 AYTSAEICLLKRFVDRGLIIQYKNSLTSPTPQLKLSVNTYFNFSINAVSLVITQD 360
QY 361 IVDYQNSLSA 370
DB 361 IVDYQNSLSA 370

RESULT 6

Q9S0F9 BORBU PRELIMINARY; PRT; 370 AA.
AC Q9S0F9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BBR11.
GN OrderedLocusNames=BBR11;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-4.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Uterback T.R., Wathey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.";
RL Nature 390:580-586 (1997).
DR EMBL: AE001577; AAF07520.1; -; Genomic_DNA.
DR TIGR: BBR11; -;
DR InterPro: IPR008505; DUF787.
DR Pfam: PF05619; DUF787; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 370 AA; 42130 MW; BE37D9DB14A1B0BD CRC64;
Query Match 96.8%; Score 1828; DB 2; Length 370;
Best Local Similarity 96.2%; Pred. No. 3e-107;
Matches 356; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
QY 1 MPQDTISVSLDSRIQASRPNNYPLLYVYKTAKIKVNDAAANYKILNLTVNNYKQIETL 60
DB 1 MPQDTISVDFDSRIQTSRPNNYPLLYVYKTAKIKVNDTASYKILNLTVNNYKQIETL 60
QY 61 EKXNGNGQDFGKEKTLTKTAMSNFFNSSESLKSADLFYKDKPEELKYLKVHRHTFV 120
DB 61 EKXNGNGQDFGKEKTLTKTAMSNFFNSSESLKSADLFYKDKPEELKYLKVHRHTFV 120
QY 121 VLINTEGNSDGLKIYKDDYDKPTSPFFVSTKEQEIKELPKDKGNTKERNIAVYS 180
DB 121 VLINTEGNSDGLKIYKDDYDKPTSPFFVSTKEQEIKELPKDKGNTKERNIAVYS 180
QY 181 NNKDNHLKFIQYHLQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
DB 181 NNKDNHLKFIQYHLQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETG 240
QY 241 LDGVPAPFEGVDLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQSGARDN 300
DB 241 LDGVPAPFEGVDLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQSGARDN 300
QY 301 AYTSAEICLLKRFVDRGLIIQYKNSLTSPTPQLKLSVNTYFNFSINAVSLVITQD 360
DB 301 AYTSAEICLLKRFVDRGLIIQYKNSLTSPTPQLKLSVNTYFNFSINAVSLVITQD 360
QY 361 IVDYQNSLSA 370
DB 361 IVDYQNSLSA 370

RESULT 7

Q9S0S1 BORBU PRELIMINARY; PRT; 362 AA.
AC Q9S0S1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BBN11.
GN OrderedLocusNames=BBN11;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid cp32-9.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=20138354; PubMed=10672174;
RA Casjens S., Palmer N., van Vugt R., Huang W.M., Stevenson B., Rosa P.,
RA Lathigra R., Sutton G.G., Peterson J.D., Dodson R.J., Haft D.H.,
RA Hickey E.K., Gwinn M.L., White O., Fraser C.M.;
RT "A bacterial genome in flux: the twelve linear and nine circular
RT extrachromosomal DNAs in an infectious isolate of the Lyme disease


```
RT spirochete Borrelia burgdorferi.;
RL Mol. Microbiol. 35:490-516(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Frazer C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M.J., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Wathey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
DR EMBL; AE001581; AAF07685.1; -; Genomic_DNA.
DR TIGR; BBN11; -;
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
DR Complete proteome; Hypothetical protein; Plasmid.
DR KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 362 AA; 41367 MW; CD76E4F607088587 CRC64;

Query Match 92.9%; Score 1755; DB 2; Length 362;
Best Local Similarity 94.2%; Pred. No. 1.2e-102;
Matches 341; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MPQDTISVSLDSRIQASRPYNYPLLYVYKTAIKVNDAAANYKILNLTNNYEQIETL 60
Db 1 MPQDTISVSLDSRIQASRPYNYPLLYVYKTAIKVNDAAANYKILNLTNNYEQIETL 60
Qy 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Db 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Qy 61 EKENGNEEDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Db 61 EKENGNEEDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Qy 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Db 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Qy 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Db 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Qy 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDVDTVIGKLTAKINFYSLN 240
Db 181 NKNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDVDTVIGKLTAKINFYSLN 240
Qy 241 LDGVPAPKEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGARDN 300
Db 241 LDGVPAPKEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGARDN 300
Qy 301 AYTSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 360
Db 301 AYTSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 360
Qy 361 IV 362
Db 361 IV 362

RESULT 8
Q5XZ68 BORGA
ID Q5XZ68 BORGA PRELIMINARY; PRT; 373 AA.
AC Q5XZ68
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN ORFNames=BG0P064;
OS Borrelia garinii PBI.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI.

Query Match 83.8%; Score 1582.5; DB 2; Length 373;
Best Local Similarity 84.1%; Pred. No. 9.1e-92;
Matches 312; Conservative 23; Mismatches 33; Indels 3; Gaps 2;

Qy 1 MPQDTISVSLDSRIQASRPYNYPLLYVYKTAIKVNDAAANYKILNLTNNYEQIETL 60
Db 1 MPQDTISVSLDSRIQASRPYNYPLLYVYKTAIKVNDAAANYKILNLTNNYEQIETL 60
Qy 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Db 61 EKONGGQDQFGKEKTLTKTAMSNFNSSESLKSADLFYKDKPEELKYLKVRHTFV 120
Qy 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 177
Db 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 177
Qy 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Db 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSTKEQEIKELFKDKGNTKERNIA 180
Qy 178 VYSNNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDVDTVIGKLTAKINFYSLN 237
Db 178 VYSNNKDLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDVDTVIGKLTAKINFYSLN 237
Qy 238 ETGLDGVPAPEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGA 297
Db 238 ETGLDGVPAPEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGA 297
Qy 241 ETGLDGVPAPEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGA 300
Db 241 ETGLDGVPAPEGVDLAGAIDEQFTHYIKNEAIIELIRIWNKNRNSKLSALQSGA 300
Qy 298 RDNATSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 357
Db 298 RDNATSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 357
Qy 301 RDNATSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 360
Db 301 RDNATSAIECLLKRFRVDRGLIIQYKNLSLTSLPTPQLKLEISVNTYFNFSINAVSLVIT 360
Qy 358 TQDIVDYQNSL 368
Db 358 TQDIVDYQNSL 371

RESULT 9
O50939 BORBU
ID O50939 BORBU PRELIMINARY; PRT; 356 AA.
AC O50939
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein BBA46.
GN OrderedLocustNames=BEA46;
OS Borrelia burgdorferi (Lyme disease spirochete).
OG Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Frazer C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Gwinn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Kerlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M.J., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Wathey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
```

RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
bursdorferi";
RL Nature 390:580-586(1997).
DR EMBL; AE000790; AAC62278.1; -; Genomic_DNA.
DR FIR; F70212; F70212.
DR TIGR; BBA46; -.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 356 AA; 40359 MW; AL1343C1F56ED474 CRC64;

Query Match 42.5%; Score 803.5; DB 2; Length 356;
Best Local Similarity 46.3%; Pred. No. 1.2e-42;
Matches 171; Conservative 58; Mismatches 117; Indels 23; Gaps 6;

QY 1 MPQDTISVSLDSRIQASRPYNYNPLLYVYTKAKIKVNDAAANYKILNLTNNYKQIETL 60
DB 1 MPKDTISVSVQNLAFNNVNYNPLLYYKS-----NVLASGHLASVTNLODLEKL 53
QY 61 EKDNNGQDQFGK-----EKTLLKTAAMSNFFNSSESLKSADLFYKDKPEELKKYLVH 115
DB 54 ENQKAQESDSKKKAEQSLKSKMSDF--GQGLKSVDFYVI-NQIKKIDFLKSN 110
QY 116 RHTFVVLINTEGNSDDGLKIYKDDYDKFKTPSIFVFSTKEQEIKELFKDKGNTKERN 175
DB 111 LHPFVVFNQAED-----VISSDFEIRKICNFIVISTKDANLPDLFGKDKSELKNI 163
QY 176 IAVYNNKDNHLKISQVYLHQAIFHAWNPGMLAATPLVDDTVIGKLTAKINFYSL 235
DB 164 IAVYA-GKQNLHLKFTAAVYLHQAIFHAWNPGMLAATPLVDDTVIGKLTAKINFYSL 222
QY 236 LNETGLDGVPAFKEGVDLAGGAIDEQTVHYIKNEAIELIRIWNKNRQNSKLSALQLS 295
DB 223 LNETGNDGLAFKEGVSLSGDPIDETFTLYIKNEAKELIRIWNKNRANSKLSGLND 282
QY 296 GARDNAYTSAIECLLKRFDVRLGIYQKMLSLTSPPTQKLELSVNIYFNSINAVSIV 355
DB 283 GNLNPEYTAGLCFFHELKQGLIVSYKEIKLINSSEGLSLSLEVALKYNDSPNSVNI 342
QY 356 ITQDIDVY 364
DB 343 ITTOEINEY 351

RESULT 10
QASIV3_BORGA PRELIMINARY; PRT; 106 AA.
AC Q5XYV3;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORFNames=BGP178;
OS Borrelia garinii PBI.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=290434;
RN [1]
RP NUCLEOTIDE SEQUENCE.

Query Match 20.2%; Score 382.5; DB 2; Length 156;
Best Local Similarity 49.1%; Pred. No. 1.5e-16;
Matches 82; Conservative 24; Mismatches 50; Indels 11; Gaps 4;

QY 82 MSNFFNSSESLKSADLFYKDKPEELKKYLVHRTFVVLINTEGNSDDGLKIYKDDY 141
DB 1 MGDFF--GQGLKSVDFYVI-NQIKKIDFLKSNLHPFVFLNESGD-----ALSSDF 50
QY 142 DKFKTPSIFVFSTKEQEIKELFKDKGNTKERNIYVYNNKDNHLKFIQSYLHQAIF 201
DB 51 EAIRKALNFIVISTKENGFLNFKGKQKGLKNIYVYSGN-ENLHLKFAAMYLHQAIF 109
QY 202 HAVNPGMPLAATPLVDDTVIGKLTAKINFYSLINTEGNDGLGVPAFK 248
DB 110 HAVNPGMILNSTFYDDSLDLSLRKANINFYSLNTEGNDGILAFK 156

RESULT 12
QASIV3_BORGA PRELIMINARY; PRT; 76 AA.
AC Q6ASII;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BGA43;
OS Borrelia garinii.
OC Plasmid lp54.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.

KW Hypothetical protein.
SQ SEQUENCE 106 AA; 11884 MW; DOA7E975F9574FB2 CRC64;

Query Match 22.3%; Score 421; DB 2; Length 106;
Best Local Similarity 83.0%; Pred. No. 3.6e-19;
Matches 83; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 MPQDTISVSLDSRIQASRPYNYNPLLYVYTKAKIKVNDAAANYKILNLTNNYKQIETL 60
DB 1 MPKDTISVSVQNLAFNNVNYNPLLYYKS-----NVLASGHLASVTNLODLEKL 53
QY 61 EKDNNGQDQFGK-----EKTLLKTAAMSNFFNSSESLKSADLFY 100
DB 61 EKDNNGSNDQLGKEQTLKTAAMSNFFNSTEESLSKSAVLFI 100

RESULT 11
QASII_BORGA PRELIMINARY; PRT; 156 AA.
AC Q6ASII;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BGA44;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.

Query Match 20.2%; Score 382.5; DB 2; Length 156;
Best Local Similarity 49.1%; Pred. No. 1.5e-16;
Matches 82; Conservative 24; Mismatches 50; Indels 11; Gaps 4;

QY 82 MSNFFNSSESLKSADLFYKDKPEELKKYLVHRTFVVLINTEGNSDDGLKIYKDDY 141
DB 1 MGDFF--GQGLKSVDFYVI-NQIKKIDFLKSNLHPFVFLNESGD-----ALSSDF 50
QY 142 DKFKTPSIFVFSTKEQEIKELFKDKGNTKERNIYVYNNKDNHLKFIQSYLHQAIF 201
DB 51 EAIRKALNFIVISTKENGFLNFKGKQKGLKNIYVYSGN-ENLHLKFAAMYLHQAIF 109
QY 202 HAVNPGMPLAATPLVDDTVIGKLTAKINFYSLINTEGNDGLGVPAFK 248
DB 110 HAVNPGMILNSTFYDDSLDLSLRKANINFYSLNTEGNDGILAFK 156

```

OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Suehnel J., Platzer M.;
RT "Comparative analysis of the Borrelia garinii genome.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000015; AAT93802.1; -; Genomic_DNA.
DR InterPro; IPR008505; DUF787.
DR Pfam; PF05619; DUF787; 1.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 76 AA; 8637 MW; 6326E818A20B5290 CRC64;

Query Match 7.4%; Score 140.5; DB 2; Length 76;
Best Local Similarity 41.9%; Pred. No. 0.12;
Matches 31; Conservative 15; Mismatches 21; Indels 7; Gaps 1;

QY 1 MPQDTISVLSLDSRIQASRPYNNPLVYTKIKVNDKAAANYKILNLTVNNYEKQIETL 60
DB 1 MPQDTISVLSLQNLVANKYNINPLLIYKS-----NTLASKHLALSVTNFEELKKL 53
QY 61 EKONGGQDQFGKE 74
DB 54 EIRKGETDSSNKK 67

RESULT 13
Q7Y202 CRYPV PRELIMINARY; PRT; 2670 AA.
ID Q7Y202_CRYPV PRELIMINARY; PRT; 2670 AA.
AC Q7Y202;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein.
GN ORFNames=IWB.772;
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bankier A.T., Spriggs H.F., Fartmann B., Konfortov B.A., Madera M.,
RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;
RT "Integrated mapping, chromosomal sequencing and sequence analysis of
RT Cryptosporidium parvum";
RL Genome Res. 0:0-0(2003).
DR EMBL; BX538353; CAD98328.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 2670 AA; 317723 MW; 670B6DF63D01A83B CRC64;

Query Match 7.4%; Score 140; DB 2; Length 2670;
Best Local Similarity 21.2%; Pred. No. 9.7;
Matches 93; Conservative 77; Mismatches 130; Indels 138; Gaps 22;

QY 3 QDTISVLSLDSRIQASRPYNNPLVYTKIKVNDKAAANYKILNLTVNNYEKQIETLEK 62
DB 1006 QESIKI-----IKSKDPNIKN--LIQKDDIKISKERN--IVNL-VNRYLSIIE---- 1050
QY 63 DNGNGQDQKGKTLTKTAMSNP-----FNSSSELSKSADLFYKPKBELKKYLVKH 115
DB 1051 DERN-----QILLILGILSRYSIRISMNLNENELLKFNLF--KDLIEK-KKVITIN 1100
QY 116 RHTFVVLN-----TEGNSDDGLK-IYKODYDKFKTPSIFVPS 154
DB 1101 SKLFGYLNDWIHFISFIINARNQIDMNDNHHFLKEWFNDNLYRYKTP--YKIFN 1158
QY 155 TKEQETK-----ELPKDKGNTKERN-----IAY 178
DB 1159 SEKILEKENVSLIFKLIKISMKYLNVNWDKNVLEIVNTRENDKSDFFYNVLV 1218
QY 179 YGNKNLHLKFSQVHQASIFHVPNGMPLAATPLVDVDTVIGKLTAKINFYSLLNE 238
DB 1031 QESIKI-----IKSKDPNIKN--LIQKDDIKISKERN--IVNL-VNRYLSIIE---- 1075

```

```
QY 63 DNGNGDQFGKKTLLTKTAMSNF-----FNSSESLSKADLFYKDKPEELKKYLVKH 115
Db 1076 DEEN-----QILLGLTSLYSIRISNNLNENELLKFNLF--KDLIEK-KKVIIN 1125
QY 116 RHTFVLIN-----TEGNSDDGLK-IYKDDYDKFKTPSIFVFS 154
Db 1126 SKLGYLINDWTHFEISFIINARMQMDMNDNHHFKLWFDNLYRYKTP--YKIFN 1183
QY 155 TKEQIK-----ELFKDKGNTKERN-----TAV 178
Db 1184 SCEKILEKNNVSLIFKLIKRNISMKYLNLVNWDKNYLEIVKNTRENDKGRDFYNVLV 1243
QY 179 YSNKNLHLKFILOYHQAISFHAVNPGMPLAATPLVDVDTVIGKLTAKINFYSLLNE 238
Db 1244 YNOSK-----VIRRIQGLDSTINRF-----NNLSENEICQMKSERIK----- 1283
QY 239 TGLDGVPAFKEGVDLAGGAIDQFTYHYIKNBAILIRIMNKNRQNSK-----LSALQL 294
Db 1284 -----EEYQRLRGLFGDLMFPKDYMNFTENDIVRNGUYNMRTSSKFLVYPPSILI 1334
QY 295 SCARDNATSAIECLIKRPVDRGLIQY-KNLSLTSLTPQKLELSVNITYN-----FSIN 350
Db 1335 ESLRSKEFENLGNPLFIPIEBISMYFSKDLALL-----VKINMTYKCBIFIID 1385
QY 351 AVSLVITTDIVDYNLSL 368
Db 1386 A-ELLITLSSEIELKSSI 1402
```

```
RESULT 15
Q7RB65 PLAYO
ID Q7RB65 PLAYO PRELIMINARY; PRT; 1057 AA.
AC Q7RB65,
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY06283;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteaux M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabai A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; ABL01002103; EAA18474.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1057 AA; 125585 MW; 9C4CB84F0A11661 CRC64;

Query Match 7.3%; Score 137; DB 2; Length 1057;
Best Local Similarity 23.5%; Pred. No. 4.8;
Matches 97; Conservative 46; Mismatches 151; Indels 118; Gaps 21;

QY 21 NYTNPLLVYTKAKIKVKNDAANY-----KILNLTWNNEKQIETLEKDN----- 64
Db 171 NYVKD-IIPKFLPKENNITQLKPNKLPFLFSLNHSLDGNDKMYQIFKKNKPNFSQLL 229
```

```
QY 65 ---GNGDQFGK--EKTLLKT--AMSNFNSSESLSKADLFYKDKPEELKKYLVKVRH 117
Db 230 SSVKNGIDGFLKYPMKTVIDTHGIYGNPLTYEYTKKEAEIYFRFNKEEDHLNPF-DIFLS 288
QY 118 TFVVLINTE-----GNSDDGLKIYKD-----DYDKFKTPSIFVFSSTKEQ 158
Db 289 FFIFISINTEIRKHLKHQVSDVDLCKLKHDPHFNFVDLIPHIDYEHIN--DVFFRFM--- 342
QY 159 EIKELFKDKGNTKERNIA-----VYSNNKDNLHLKFILOYHQAISFHAVNPGM 209
Db 343 -----NTTKTHNILPLMLFRLFFHPKND-----LFTYSNVINAVEKY-- 381
QY 210 PLAATPLVDDT--VIGKLTAKINFYSLLNETGLDGVPAFKEGVD--LAGGAIDQFTVH 265
Db 382 -FTSEIFDSDTKKMIGKLA-----YSLMARLSFITVSYLEPKIDDDVIDGIDIDLMNLFH 434
QY 266 YIKNEAIIEL-----IRWNKNRQNSKLSALQ-----LSGARDNATYSAIEC 308
Db 435 NVINKMVVTLCKEDKKLIEFPFVPISMVKNKMKFKLDALKIYNBFTITSLPDQDKESAINF 494
QY 309 LLKRP--VDRGLI---IQYKNLSLTSLP-----TPQLKLELSVNITYNFSIN 350
Db 495 LQTDPPKVVVTTLSLDALDFVNRHLEKGPDTYHSDEYVKLETFLFKINYQISSN 546

Search completed: January 24, 2006, 19:57:00
Job time : 81.505 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:30:59 ; Search time 22.8097 Seconds
(without alignments)
1341.095 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPQDTISVLLDSRIQASRP.....AVSLVITQDIVDYQNSLSA 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5 COMB.pdp:*

2: /cgn2_6/ptodata/1/iaa/6 COMB.pdp:*

3: /cgn2_6/ptodata/1/iaa/H COMB.pdp:*

4: /cgn2_6/ptodata/1/iaa/PCRTUS COMB.pdp:*

5: /cgn2_6/ptodata/1/iaa/RE COMB.pdp:*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.5	7.0	1155	2	US-09-710-279-1780
2	130.5	6.9	1742	2	US-09-386-962C-4
3	130.5	6.9	1742	2	US-09-386-959-4
4	119	6.3	612	2	US-09-538-092-775
5	116	6.1	524	2	US-09-543-681A-4787
6	112.5	6.0	340	2	US-09-949-016-10217
7	112.5	6.0	382	1	US-08-477-451-28
8	111	5.9	1087	2	US-09-830-230A-322
9	111	5.9	1119	2	US-09-830-230A-321
10	111	5.9	1241	2	US-09-248-796A-18397
11	110	5.8	1078	2	US-09-248-796A-20284
12	109	5.8	1164	2	US-08-923-992A-10
13	107	5.7	429	2	US-09-134-000C-5630
14	107	5.7	544	2	US-09-710-279-38
15	105	5.6	1060	2	US-09-248-796A-15476
16	105	5.6	1098	2	US-08-923-992A-8
17	104.5	5.5	400	2	US-09-248-796A-15785
18	104.5	5.5	458	2	US-09-830-230A-30
19	104.5	5.5	481	2	US-09-830-230A-29
20	104.5	5.5	561	2	US-09-248-796A-24349
21	104	5.5	1164	2	US-08-923-992A-2
22	103.5	5.5	491	2	US-09-029-267-2
23	103.5	5.5	677	2	US-09-543-681A-5460
24	103.5	5.5	691	2	US-09-538-092-421
25	103.5	5.5	3878	1	US-09-914-259-11
26	103	5.5	1781	1	US-08-477-451-11
27	102	5.4	488	2	US-09-248-796A-18820

28	102	5.4	3135	1	US-08-323-170B-2	Sequence 2, Appli
29	102	5.4	3135	2	US-08-954-441-2	Sequence 2, Appli
30	101.5	5.4	439	2	US-09-538-092-1216	Sequence 1216, Ap
31	101.5	5.4	736	2	US-09-583-110-4555	Sequence 4555, Ap
32	101.5	5.4	741	2	US-09-107-433-4401	Sequence 4401, Ap
33	101.5	5.4	773	2	US-09-841-786-6	Sequence 6, Appli
34	101.5	5.4	3241	2	US-09-841-786-1	Sequence 1, Appli
35	101	5.3	820	2	US-09-248-796A-19901	Sequence 19901, A
36	101	5.3	912	1	US-08-951-871-2	Sequence 2, Appli
37	101	5.3	972	2	US-09-248-796A-20234	Sequence 20234, A
38	100.5	5.3	2184	2	US-09-417-485D-6	Sequence 6, Appli
39	100.5	5.3	2710	1	US-08-480-604A-6	Sequence 6, Appli
40	100.5	5.3	2710	2	US-08-405-496A-6	Sequence 6, Appli
41	100.5	5.3	2710	2	US-08-915-136-6	Sequence 6, Appli
42	100.5	5.3	2710	2	US-08-957-310-6	Sequence 6, Appli
43	100.5	5.3	2710	2	US-10-011-366-6	Sequence 6, Appli
44	100.5	5.3	2710	2	US-09-084-517-6	Sequence 6, Appli
45	100	5.3	717	2	US-09-107-532A-6277	Sequence 6277, Ap

ALIGNMENTS

RESULT 1

US-09-710-279-1780
; Sequence 1780, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR FILING DATE: 2000-11-09
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-1780

Query Match	7.0%	Score 132.5;	DB 2;	Length 1155;
Best Local Similarity	22.3%	Pred. No. 0.0048;		
Matches	96;	Conservative	58;	Mismatches 169;
				Indels 107;
				Gaps 21;
Qy	18	SRPNYPLLVYKTAIKVVK-DAANYKILMLTVNVEKQIETLEKONGQDOFGKEKT	76	
Db	113	SKPRANEAVTNSTKPKTTEAPTVEESIAETPKTSTTQODSTKKNPSLKDNLNLSST	172	
Qy	77	LLKTAMNFFNNSSESLSADLFYK-----DKPEELKKYLKVHRHFFVLINTEGNS	130	
Db	173	TSKES-----KTDEHSTKQACMTKSNLNDTNDSTPQSEK-----TSSQANDSTDN	220	
Qy	131	D-----DGLKIVKDDYD-----KFKTPTSIFFVFSTKEQ-----IKE	162	
Db	221	SAPSKQSDSKSEKQVYKTFNDPTQVHTTKLTPSISTDSSVNDKQDYTRSAVAS	280	
Qy	163	LPKDKGNTKEKRNATVYNNKNDLHLKFPIS-QYLHQASIFHAV-----NP-YGM--PLAA	213	
Db	281	LGVDNTE-----AITNAVNDLNDLKAASREQINEALIAALKKDFSNPDYGVDTPLAL	335	
Qy	214	TPLVDDTVIGKLTAKINFYSLINETGL-----DCVPAPKEGVDL-----AGGAI----	258	
Db	336	NTSQSKSPHSKASPRMNLMSLAAEPNSGKNVDKVKITNPTLSLNKNHANNVWPTS	395	
Qy	259	DEQFTY--HYTKNEAI-----IELIRIWNKNRQNSKLSALQLSGAR	298	
Db	396	NEQFNLKANYELDDSIKEGDTFTIKYQYIRPGGLELPAI---KTQLRSKQGSIVANGVY	452	


```
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPL070W
US-09-538-092-775

Query Match          6.3%; Score 119; DB 2; Length 612;
Best Local Similarity 21.3%; Pred. No. 0.032;
Matches 86; Conservative 58; Mismatches 138; Indels 122; Gaps 20;

Qy 19 RPNYPLVYKTAKI-----KVNDAANYKILNLTNN--YEKQIETLEKNGNGQDQFG 72
Db 102 QPKVRLVQLSLFQSFYIKDKASFQHVSSANNNGY-----PSGGSSSFL 152

Qy 73 KEKILLTAMSNFNSSSES-----LKSADLIYKD-----KPELKKYLK 113
Db 153 AAKETLSSGLSGIFGRSSSSGNSILMRPRSSSLFSNESISNSTNATQMLSPKEIKQLK 212

Qy 114 VHRHFFVLIINTEGNSDGLKIYK----DDYDKETPSIFFVSTKEQIKELFKDGN 169
Db 213 I-----NELNMKIEKYMELCERDVPK--KILIV----- 239

Qy 170 TEKERNIAVYNNKDNHLKFIQYLHOASIFHAVNPGMPLAAT---PLVDDTV--IGK 224
Db 240 -----GTSVSSPNK---MKTFKP--HQIQTFKVGNLFNSVEFTYNNKLNKILCLSK 288

Qy 225 LRTA-KINFYSLN-ETGLDGPAPKEGVDLAGAIDQFTYHYI---KNEAIELIRI 278
Db 289 LSTWKNILKFLSLNNGIDPEKPEIKDIL-----YEFTYHSISPCEKIKALLKHEI 343

Qy 279 WKNKNRQNSKLSALQSGARDNAYTSAIECLLKRFDVDRGL-----IIQYKNLSLTS 330
Db 344 MTYSQEMS-----NDDYLSLIYIITIVPRDIFLNAEFILFRYKFKKLVE 391

Qy 331 PTPQLKLELV-----NITVNSINAVSLVITTDIVDYQNSLSA 370
Db 392 SFALTNLNAVFEGLTDFNSNELQDKLTVNESKILENSISS 435

RESULT 5
US-09-543-681A-4787
; Sequence 4787, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4787
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4787

Query Match          6.1%; Score 116; DB 2; Length 524;
Best Local Similarity 19.7%; Pred. No. 0.048;
Matches 71; Conservative 51; Mismatches 144; Indels 94; Gaps 12;

Qy 1 MPQDTISVLLDSRTOASRPYNNPLVYKTAKIKVNDAANYKILNLTNNYKQIE-- 58
Db 245 LENDNINLNDNINLNDNINLNDNINLNDNINLNDNINLNDNINLNDNINLNDN 304

Qy 59 --TLEKNGNGDQFGKEKTLTKTAMSN--PNSSESLKSADL-----FYKDKPEEL 108
Db 305 NINLENDNAN-----LNKIKALIKINRNTYIFNGKEIKSQODKAMKGFETLTKDIS 360

; LENGTH: 612
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YPL070W
US-09-538-092-775

Query Match          6.3%; Score 119; DB 2; Length 612;
Best Local Similarity 21.3%; Pred. No. 0.032;
Matches 86; Conservative 58; Mismatches 138; Indels 122; Gaps 20;

Qy 19 RPNYPLVYKTAKI-----KVNDAANYKILNLTNN--YEKQIETLEKNGNGQDQFG 72
Db 102 QPKVRLVQLSLFQSFYIKDKASFQHVSSANNNGY-----PSGGSSSFL 152

Qy 73 KEKILLTAMSNFNSSSES-----LKSADLIYKD-----KPELKKYLK 113
Db 153 AAKETLSSGLSGIFGRSSSSGNSILMRPRSSSLFSNESISNSTNATQMLSPKEIKQLK 212

Qy 114 VHRHFFVLIINTEGNSDGLKIYK----DDYDKETPSIFFVSTKEQIKELFKDGN 169
Db 213 I-----NELNMKIEKYMELCERDVPK--KILIV----- 239

Qy 170 TEKERNIAVYNNKDNHLKFIQYLHOASIFHAVNPGMPLAAT---PLVDDTV--IGK 224
Db 240 -----GTSVSSPNK---MKTFKP--HQIQTFKVGNLFNSVEFTYNNKLNKILCLSK 288

Qy 225 LRTA-KINFYSLN-ETGLDGPAPKEGVDLAGAIDQFTYHYI---KNEAIELIRI 278
Db 289 LSTWKNILKFLSLNNGIDPEKPEIKDIL-----YEFTYHSISPCEKIKALLKHEI 343

Qy 279 WKNKNRQNSKLSALQSGARDNAYTSAIECLLKRFDVDRGL-----IIQYKNLSLTS 330
Db 344 MTYSQEMS-----NDDYLSLIYIITIVPRDIFLNAEFILFRYKFKKLVE 391

Qy 331 PTPQLKLELV-----NITVNSINAVSLVITTDIVDYQNSLSA 370
Db 392 SFALTNLNAVFEGLTDFNSNELQDKLTVNESKILENSISS 435

RESULT 6
US-09-949-016-10217
; Sequence 10217, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10217
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10217

Query Match          6.0%; Score 112.5; DB 2; Length 340;
Best Local Similarity 20.3%; Pred. No. 0.052;
Matches 60; Conservative 45; Mismatches 101; Indels 89; Gaps 13;

Qy 25 PLLVYKTAKIKVNDAANYKILNLTNNYKQIETLEKNGNGDQFGKEKTLTKTAMSN 84
Db 102 PVFIRKDRVTQ-----EREAEAL-----KQKELEQEA-- 131

Qy 85 FNSSESLKSADLIYKDKPEELKYLKVRHRTFVVL--INTGNSDDGLKIYKDYD 142
Db 132 --RMAEERRK---YTLKIVEETKKELENKRSIALDALNTDDENDE-----EYEE 178

Qy 143 KFKTPSIFVSTKEQIKELFKDGNTEKERNIA-----VYSNNKDNHLK 189
Db 179 AWKVELKRIKRDREDR-EALEKEKAEIERMNLTEERRAELRANGKVIITNAVKGYK 237

Qy 190 FISQYLHOASIF--HAVNPGMPLAATPLVD---DTVIGKLRTAKINFYSLNLTGLDV 244
Db 238 FLQKYHKGAFMEDDEBEVYKRDPSAPTELDHFNKLTILPKVNVQK-NF-----GRSGR 289

Qy 245 PAFKEGVDLAGAIDQFTYHYIKNEAIELIRIWNKNRQNSKLSALQSGARD 299
Db 290 TKYTHLVDDTTSFDS-----AWQBSAQNTKFKQKAAGVRD 327

RESULT 7
US-08-477-451-28
; Sequence 28, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
```

; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 28:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 382 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-477-451-28

Query Match 6.0%; Score 112.5; DB 1; Length 382;
 Best Local Similarity 24.8%; Pred. No. 0.062;
 Matches 81; Conservative 40; Mismatches 117; Indels 89; Gaps 18;
 QY 27 LVYTKAKIKVNDKANYKILNLTNNYKQIETLEKNGNGQDQFGKEKTLTKTAMSNFF 86
 Db 67 LIFRQKIKELMD-HRAKYSLDNKYKKEALEKETR-----GK-----ILTKASKAY 115
 QY 87 NSSESLKSADLFIYKDKPEELKYLKVRHTFVVLINTEGNSDD--GLKIYKDDYDKF 144
 Db 116 GLEQAL-----KDNP-----LYKLLPNPYAVLNQETTKEDKERLSY---YQV 160
 QY 145 KPSIF-----FVSTKEQBIKELFKDKGNTKERNIAVYSNNKDL- 186
 Db 161 KTSSIFETKTATTKDQAQLQMGVSLDEQNK--ASRLALSYKQAIIEEYSSNNISNL 218
 QY 187 ---HLKFTSOYLHQAISIHANVPYGMPL--AATPLVDDTVIGKLTAKINFYSLNETGL 241
 Db 219 SRKELNDIDYQLQER-----NKFOSKAKDIAQKATNTLIFNSERLAFSAIDKINEKYL 273
 QY 242 DGVPAF-----KEGVLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQL 294
 Db 274 KYEAFSNLLKNVDDVEL--NTLFKNFT-----NQKL-----SFAQKQLCLLVL 317
 QY 295 SGARDNAVTSATIECLLR-----FVD 315
 Db 318 DSF--NFDTSKKSILKKTNEYNI FVD 342

RESULT 8
 US-09-830-230A-322
 ; Sequence 322, Application US/09830230A
 ; Patent No. 6902893
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc.
 ; TITLE OF INVENTION: Lyme Disease Vaccines
 ; FILE REFERENCE: PB481US
 ; CURRENT APPLICATION NUMBER: US/09/830,230A

; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US98/12718
 ; PRIOR FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: 60/057,483
 ; PRIOR FILING DATE: 1997-09-03
 ; PRIOR APPLICATION NUMBER: 60/053,344
 ; PRIOR FILING DATE: 1997-07-22
 ; PRIOR APPLICATION NUMBER: 60/053,377
 ; PRIOR FILING DATE: 1997-07-22
 ; PRIOR APPLICATION NUMBER: 60/050,359
 ; PRIOR FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 756
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 322
 ; LENGTH: 1087
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (541)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (595)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (703)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-09-830-230A-322

Query Match 5.9%; Score 111; DB 2; Length 1087;
 Best Local Similarity 22.6%; Pred. No. 0.43;
 Matches 74; Conservative 50; Mismatches 122; Indels 82; Gaps 15;
 QY 37 NKDAANYK--ILNLTNNYKQIETLEKNGNGQDQFGKEKTLTKTAMSN-----FFNS 88
 Db 741 NTQEDHYKLGIRFKLKYHESIESF-----DQTIKLPKHKKALHNGKIALMLNKK 792
 QY 89 SEESLSKADLFIYKDKPEELKYLKVRHTFVVLINTEGNSDDGLKIYKDDYDKFPTPS 148
 Db 793 NKKAIESFEKAIQIDKNGYGTAYQK-----GIAEKNQDMQQAFAFKAYNLDKNPN 845
 QY 149 IFF---VFST-----KEQIKELFKDKGNTKERNIAVYS-----NNKDLHLKFI- 191
 Db 846 YALKAGIVSNLGNFKQSEEVINFF--NANAKKPEIAIYNLSIAKPENKLEESLETIN 903
 QY 192 -----SOYLH-QASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNETGL 241
 Db 904 KAILNPKSEYLYLKASINLKENY-----QNAISLYSLVIEKNP 944
 QY 242 DGVPAFKEGVLAGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQSGARDNA 301
 Db 945 ENTSAY---INLA-----KAYEKGNSQALSTLEKIINKNK--LALNNLGIYKKEKN 994
 QY 302 YTSATIECLLRKFRDGLIIQVKNLSLT 329
 Db 995 YOKAIEIFEKAIINSIDIEAKY-NLATTL 1021

RESULT 9
 US-09-830-230A-321
 ; Sequence 321, Application US/09830230A
 ; Patent No. 6902893
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc.
 ; TITLE OF INVENTION: Lyme Disease Vaccines
 ; FILE REFERENCE: PB481US
 ; CURRENT APPLICATION NUMBER: US/09/830,230A
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US98/12718
 ; PRIOR FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: 60/057,483

; PRIOR FILING DATE: 1997-09-03
 ; PRIOR APPLICATION NUMBER: 60/053,344
 ; PRIOR FILING DATE: 1997-07-22
 ; PRIOR APPLICATION NUMBER: 60/053,377
 ; PRIOR FILING DATE: 1997-07-22
 ; PRIOR APPLICATION NUMBER: 60/050,359
 ; PRIOR FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 756
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 321
 ; LENGTH: 1119
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: SITE
 ; LOCATION: (573)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (627)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (735)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-09-830-230A-321

Query Match 5.9%; Score 111; DB 2; Length 1119;
 Best Local Similarity 22.6%; Pred. No. 0.45;
 Matches 74; Conservative 50; Mismatches 122; Indels 82; Gaps 15;
 QY 37 NKDAANYK--ILNLTNNYKQIETLEKONGNGQDQFQKTKLLKTAMSN-----FPNS 88
 DB 773 NTQEDHYKGIIRFKLKKYHSIESF-----DQTKLPKPKKHLNKGIALMLNK 824
 QY 89 SEESLSADLFYKDKPEELKYLKVRHTFVVLNTEGSDSDGLKIYKDDYDKFTPS 148
 DB 825 NKKAIESPEKAIQIDKNYGTAYQK-----GIAEERKNGMQQAFKFNAYNLDKPN 877
 QY 149 IFF--VFST-----KEQIKELPKDKGNTKERNIAVYS-----NKNLHLKEI- 191
 DB 878 YALKAGIVSNLGNFKQSEYVLPF--NANAKPNEIALYLSIAKFNKLESELEIN 935
 QY 192 -----SOYLH-QASIFHAVNPGMPLAATPLVDVTGKLTAKINFYSLNLTGL 241
 DB 936 KAILDNPEKSEYLYLKASINLKKENY-----QNAISLYSLVIEKNP 976
 QY 242 DGVPAFKEGVDLAGAIDEQFTYHYKNEAIIELIRIWNKNRQNSKLSALQSGARNA 301
 DB 977 ENTSAY---INLA-----KAYEKGNSQASITLEKIKNKNK--LALNNLGIYKKEKN 1026
 QY 302 YTSAECLLKRFVDRGLIIQYKNLSLT 329
 DB 1027 YOKAIEIPEKALINSIDIEAKY-NLATTL 1053

RESULT 10
 US-09-248-796A-18397
 ; Sequence 18397, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstein et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 18397

; LENGTH: 1241
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-248-796A-18397
 Query Match 5.9%; Score 111; DB 2; Length 1241;
 Best Local Similarity 20.9%; Pred. No. 0.53;
 Matches 87; Conservative 58; Mismatches 162; Indels 110; Gaps 18;
 QY 4 DTISVSLDSRIQASRPNY-----YNPLL--VYKTAKI--KVNKDAANYKIL----- 46
 DB 792 DNTSTTTNTKYIRAPKLKYESMFLIYNDKLLPIYNSATILFLKIELVESNKWYLLKFGNL 851
 QY 47 -NLTVNNEYKQIETLEKONGNGQDQF-----GREKTIILKTAMSNFFNSSESL 93
 DB 852 RNLQIKNTSEDIQKHLNIDEANQYFEIDNTVDLSTVINEPKTLLNLNIFNTLKNLSLI 911
 QY 94 KSADLFYKDKPEELKYLKVRHTFVVLNTEGSDSDGLKI--YKDDYDKFTPSFFV 152
 DB 912 KILD-----QNKSNVTVL-----DMSMDNITINIKKYNDNDKLLIILK 951
 QY 153 FSTKEQIEIKELPKDKGNTKERNIAVYSNN--KDLHLKFIISQYLHQASI----- 200
 DB 952 LPEQATDSIQLLMKNGSTTTTDELDLNNNILEFELILKFLNQYLRESKLNHNRRQIISI 1011
 QY 201 -----FHAVNPGMPLAATPLVDVTGKLTAKINFYSLNLTGLDGVPAFKEGVDLA 254
 DB 1012 IKIIQYLTEINPI---LQSTKAINQQLAELKR-----INSTNNNNNAKRRIILKLS 1059
 QY 255 GGAIDEQFTYHYKNEAIIELIRIWNKNRQNSK-----LSAL--QLSGARNA 301
 DB 1060 NGLYKFLVNLNII--SLTHLQLVFFMNSNTGSLNKKIQDKIMINLSLIFKDFRFSKPNGF 1118
 QY 302 YTSAECLLKRFVDRGLI---IQYKNL-----SLTSPQLKLELSVNITYN 346
 DB 1119 QDSTNGHLIKISFKDSLINENLKEFNLFKINELLITKSKQIK---SLNQTEIN 1172

RESULT 11
 US-09-248-796A-20284
 ; Sequence 20284, Application US/09248796A
 ; Patent No. 6747137
 ; GENERAL INFORMATION:
 ; APPLICANT: Keith Weinstein et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.132
 ; CURRENT APPLICATION NUMBER: US/09/248,796A
 ; CURRENT FILING DATE: 1999-02-12
 ; PRIOR APPLICATION NUMBER: US 60/074,725
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: US 60/096,409
 ; PRIOR FILING DATE: 1998-08-13
 ; NUMBER OF SEQ ID NOS: 28208
 ; SEQ ID NO 20284
 ; LENGTH: 1078
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-09-248-796A-20284

Query Match 5.8%; Score 110; DB 2; Length 1078;
 Best Local Similarity 21.6%; Pred. No. 0.52;
 Matches 79; Conservative 59; Mismatches 140; Indels 88; Gaps 14;
 QY 44 KILNLTNNYKQIETLEKONGNGQDQFQKTKLLKTAMSNFFNSSESLKASDLFIYKD 103
 DB 367 KVLNLT-----KELQT-EKENAESNDKELNEKI--EKLTLNLTSLKLETKL-----ED 409
 QY 104 KPEELKYLKVRH--HTFVVLINT-----EGDSDDG 133
 DB 410 KEQELAKIQEDHKLNLNEKFLVTNLSNKKASKKEFETASQYKELQALKKGNTSST 469
 QY 134 LKIYKDDYDKPK-----TPSIFV-----FSTKEQIEIKELFKDKGNT 170

Db 470 LKQLEKLDSTEQAKKLEGGDGINNTRDLFLHKLKSEAFTQIKREREFKNLTVELENT 529
QY 171 EKERNIAVYNNKNLHLK----FISYLHQASIFHAVNPGM--PLAATPLVDDTVIGK 224
Db 530 KADYELQINLNKSNNEFKQKINELUSKKIESLTEDNKSNAKOLEBKJLDRDTEENHLMOK 589
QY 225 LRTAKINFYSL---LNETGLDGPAPKEGVDLAGGAID----EQPTYHYIKNEAIEILIR 277
Db 590 LRSASVAYNDLKAKSESEETVKA-KEELETLTSDKIDNLEKELKEQSQKNELEGGQLQN 648
QY 278 INWKNRQNSKLSALQSGARDNAVTSIAIECLLKRFPVDRGLIIQVKNLSLTLSPPPQLKL 337
Db 649 ITDSTNEKFELEDELKSTIKSNKEISSONSELIOKLEK----TEKDLQAKDEEIDKLKA 704
QY 338 ELSVNI 343
Db 705 ETKSNI 710

RESULT 12

US-08-923-992A-10
; Sequence 10, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1164 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-923-992A-10

Query Match 5.8%; Score 109; DB 2; Length 1164;
Best Local Similarity 22.6%; Pred. No. 0.73;
Matches 83; Conservative 47; Mismatches 134; Indels 104; Gaps 18;
QY 55 KOIETLEKONGQDFGKEKTL-----LKTAMSNF-----FNSSEESLKSADLFIYKDK 104
Db 165 KOHEVEKDKKAKQ-----QKTLKQSDTKVDLSNIDKELNHQKSOVETMAEOLGTNEDK 219
QY 105 PEELKYLKVRHRTFVVLINTEGDSDDGLKIYKDDYDKFKTPSFFVFSTK---EQRIK 161

Db 220 DSKLKIETDIRKQA-----QQADKEDAEVKKVEELGKLF-----STKAGLDQBIQ 266
QY 162 ELFKDKGKTEKERNIAV---YSNNKNLHLKFI-----SOYLHOASIFH 202
Db 267 BHVK-KETTSEENTQKDEHYFPNSLQNLQAQSLSELBKATTEQATQVKNQFLENAQKLK 325
QY 203 AVNPGMPLAATPLVDDTVIGKLTAKINFYSLNETGLDGPAPKEGVDLAGGAIDRQF 262
Db 326 EIQP-----LIKETNVKLYKAMSESE-----QVEKQL 353
QY 263 TYHYIKN-EAII-----ELIRIWNKNRQNSKLSAL-QLSGARDNAVTSIAIECLLKRFPVDR 316
Db 354 KHNSQANLEDLVAKSKETVREYEGKLNQSKNLPKLEEBEAHSLKQKQVEDFRKFKPTS 413
QY 317 GLIIQYKNL--SLTLSPPPQLKLELSV---NIT-----YNFSINAVSLVITTDIVD-- 363
Db 414 EQVTPKRVKRDLAANENNOQKIELTVSPENITVYEGEDLKFTLTAKSDSKTTLDFSDLL 473
QY 364 --YQNSLS 369
Db 474 TKYNPSVS 481

RESULT 13

US-09-134-000C-5630
; Sequence 5630, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5630
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5630

Query Match 5.7%; Score 107; DB 2; Length 429;
Best Local Similarity 20.1%; Pred. No. 0.24;
Matches 71; Conservative 46; Mismatches 128; Indels 108; Gaps 11;
QY 6 ISVSLDSRIQASRPNYNPLLVYKTAKIKVKNDAANYKIILNLTWNVEKQIETLEKONG 65
Db 153 ITIALIDSGIDRLHPNLQDN-----NLRKNYVNDIELDEYGHG 191
QY 66 -----NGDQFGKEKTLTKTAMS-----NFFNSEESLSKAD 97
Db 192 TVAGVIDTIAPRVNLNSYKVMGTGTD--GNSINMLKAIVDATNDQVDIINVSLSGYKNME 249
QY 98 LFIYKDPPEELKYLKVRHRTFVVLINTEGDSDDGLKIYKDDYDKFKTPSFFVFSTKE 157
Db 250 IDDERFTVEAFRAKAVNYARKNNILIVASAGNESRD-----ISTGNE 290
QY 158 QEIK---ELFKDKGNTKERNIAVYNNKNLHLKFIISQYLHQASIFHAVNPGMPLAAT 214
Db 291 KHIFGGLESVITVGATKSGDIADYSNYGSNV-----SIYGPAGGYGDNKYIT 338
QY 215 PLVDDTVIGKLTAKINFYSLNETGLDGPAPKEGVDLAGGAIDEQPTYHYIKNEAIE 274
Db 339 GQID-----AREMMVTYPTSLVPLGKAADFPDGYTLSPGT-----SLATPEVSAA 385
QY 275 LIRIWNKNRQNSKLSALQSGARDNAVTSIAIECLLKRFPVDRGLIIQVKNLSL 327
Db 386 LAAIMSK-NVDNSKSDNSEVLNTLPENA-----DSFIDKNSMLKYKEVRI 428

RESULT 14

US-09-710-279-38
; Sequence 38, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-38

Query Match 5.7%; Score 107; DB 2; Length 544;
Best Local Similarity 22.4%; Pred. No. 0.35;
Matches 71; Conservative 51; Mismatches 95; Indels 100; Gaps 17;
QY 18 SRPNYNPL--VYTKAKVKNDAANYKILNLTNNYKQIETLEKONGQD-----Q 70
DB 2 SMENHIERLINHVEKTIK-----EYAFSLGKSNITKAKVKLKKPNLRRDITKBIQ 55
QY 71 FGKKEKTLTKAMSNFFNSSEESLSADLFYKDKPEEL-----KYL----- 112
DB 56 KFRQKTGAPSPWKIDIVTER-----EVLFDKVDKDELQTRNVIDFGIALDQYWNLSF 110
QY 113 ---KVHRHTFVVLINTEG-----DMSDDGLKTY-----KDDYD-----KFKTPS 148
DB 111 LPEEINTNAFIKPVKTGDKTLILSEQINNLYRKVYTHHKKKFAVDYFVENKEVIFKFKTG 170
QY 149 IPFVSTKEQIEKELFKDKGNTKEKRNIAVYNNKONLHLKFIQSLHQAASIFHAVNPG 208
DB 171 ----FILDEQKIYEL-HDEGYKKGLRKV-----DYLH-KEIDQLIESGTYF-----LG 212
QY 209 MPLAAT-----PLVDDTVIGKLTAKINFYSLINETG-----LDGVPAPKEGVDLAG 255
DB 213 NMLSDTGRYQYGYFPHFD-----XEIFNYILRHASSTYALIEGLDYLIGEDLTIVE 263
QY 256 GAIDEQFTVHYIKNEAI 272
DB 264 KAINVVIENFYDNEGV 280

RESULT 15

US-09-248-796A-15476
; Sequence 15476, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15476
; LENGTH: 1060
; TYPE: PRT
; ORGANISM: Candida albicans

US-09-248-796A-15476

Query Match 5.6%; Score 105; DB 2; Length 1060;
Best Local Similarity 20.0%; Pred. No. 1.5;
Matches 79; Conservative 72; Mismatches 150; Indels 94; Gaps 18;
QY 10 LLSRIQASRPNNYNPLVYTKAKVKNDAANYKILNLTNNYKQIETLEKONGQD 69
DB 106 VFNEKVGSSPSWN--ILLY-----VPSDAKQKQSTIIKMSHYDKLLKDFSNEGGKELA 157
QY 70 QFGKEKTLTKAMSNFFNSSEESLSADLFYKDKPEELKYL-----KVHRHT 118
DB 158 ALFAESTSPASV-----GAQNSKSTGYCFKFKQHEFELNEFLVQIKNLLGFTFNQKYHL 212
QY 119 FVVLINTEGSDGLKIYKDDYD-----KFKTPSIFVFTSTKEQIEKELFKDKGNT 170
DB 213 NLDLITTSGDHENSLSLTKYVATYNLAELFYDMKLFNDCINF-FNRLSQLNTLVENPNL 271
QY 171 ---EKERNIAVYNNKONLHLKFIQ-----QYLHQAASIFHAVNPGMPL-----AAT--PL 216
DB 272 FIYKVDLPAPKVFNSKFD--FKKFYNNKCOHMDVNSFTNNVLFELKCFIFFRQASTLEML 329
QY 217 VDDTVIGKLTAKINFYSLINETG---DGVPAFKEGVDLAGGAIDEQFTVHYIK----- 268
DB 330 VKNLNNSISLAELQISKLRLNLVLFNLDLLQVFQNEQALI-----EFYSIIEYFLNL 383
QY 269 ---NEAIIELIRIW-----NKN-----RQNSKLSALQLSGARDNAYTSAIECLLKR 312
DB 384 DIVNKLIEQATKLYEAEPANNNSYQLKRFESRGELKLLQRSSLIKARENSIBI----- 439
QY 313 FVDRGLIIQYKNLSI-----TISPTPQLKLEL 339
DB 440 ---KGLDQVPEDVSLDBEEGPKQSKDSTEQVKLDL 471

Search completed: January 24, 2006, 20:03:01
Job time : 23.8097 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:57:15 ; Search time 65.0296 Seconds
(without alignments)
2377.329 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPQDTISVSLDSRIQASRP.....AVSLVTTQDIVQNSLSA 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1889	100.0	370	5	US-10-688-058-20 Sequence 20, Appl
2	1837	97.2	370	5	US-10-688-058-64 Sequence 64, Appl
3	788	41.7	160	5	US-10-688-058-62 Sequence 62, Appl
4	329	17.4	65	5	US-10-688-058-18 Sequence 18, Appl
5	137.5	7.3	1633	4	US-10-282-122A-70437 Sequence 70437, A
6	130.5	6.9	1742	4	US-10-615-383-4 Sequence 4, Appl
7	130.5	6.9	1742	4	US-10-690-184-4 Sequence 4, Appl
8	130.5	6.9	1742	4	US-10-689-082-4 Sequence 4, Appl
9	129.5	6.9	1877	5	US-10-732-923-22524 Sequence 22524, A
10	129.5	6.9	1877	5	US-10-732-923-22547 Sequence 22547, A
11	120.5	6.4	575	4	US-10-724-972A-7483 Sequence 7483, Ap
12	120.5	6.4	598	4	US-10-282-122A-54216 Sequence 54216, A
13	119	6.3	1178	4	US-10-282-122A-72168 Sequence 72168, A
14	119	6.3	1331	4	US-10-282-122A-63427 Sequence 63427, A
15	117.5	6.2	5005	4	US-10-282-122A-76871 Sequence 76871, A
16	117	6.2	806	5	US-10-732-923-23131 Sequence 23131, A
17	117	6.2	1156	4	US-10-369-493-43 Sequence 43, Appl
18	116.5	6.2	1191	4	US-10-282-122A-52048 Sequence 52048, A
19	116.5	6.2	6761	5	US-10-732-923-15035 Sequence 15035, A
20	115	6.1	590	4	US-10-282-122A-52497 Sequence 52497, A
21	113.5	6.0	596	4	US-10-282-122A-52435 Sequence 52435, A
22	113	6.0	1805	3	US-09-820-843A-73 Sequence 73, Appl
23	113	6.0	1805	4	US-10-282-122A-63513 Sequence 63513, A
24	113	6.0	1805	5	US-10-732-923-3303 Sequence 3303, Ap
25	112.5	6.0	1265	4	US-10-032-585-7168 Sequence 7168, Ap
26	112	5.9	1394	4	US-10-437-963-183281 Sequence 183281, A
27	112	5.9	1881	4	US-10-032-585-7646 Sequence 7646, Ap

ALIGNMENTS

RESULT 1

US-10-688-058-20
; Sequence 20, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:0030US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-10-688-058-20

Query Match	100.0%	Score	1889	DB	5	Length	370
Best Local Similarity	100.0%	Pred. No.	5e-144				
Matches	370	Conservative	0	Mismatches	0	Indels	0
Gaps	0						

Qy	1	MPQDTISVSLDSRIQASRP	NYNPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL	60
Db	1	MPQDTISVSLDSRIQASRP	NYNPLLVYKTAKIKVNDAAANYKILNLTNNYKQIETL	60

Qy	61	EKDNNGODQFGKTKLLKTAMSNFNSSESLSKADLFYKDPPEELKKYLKVRHFTV	120
Db	61	EKDNNGODQFGKTKLLKTAMSNFNSSESLSKADLFYKDPPEELKKYLKVRHFTV	120

Qy	121	VLINTEGNSDDGLKYKDDYDKETPSIPFVFTSKQEIKELPKDKGNTKERNIAVYS	180
Db	121	VLINTEGNSDDGLKYKDDYDKETPSIPFVFTSKQEIKELPKDKGNTKERNIAVYS	180

Qy	181	NNKDNHLKFTISQYLHQASIFHAVNPGMPLAATPLVDVDTVIGKLTAKINFYSLNBTG	240
Db	181	NNKDNHLKFTISQYLHQASIFHAVNPGMPLAATPLVDVDTVIGKLTAKINFYSLNBTG	240

Qy	241	LDGVPAPKEGVDLAGGAIDEQFTYHIKNEAIIELIRIWNKNRNSKLSALQLSGARDN	300
Db	241	LDGVPAPKEGVDLAGGAIDEQFTYHIKNEAIIELIRIWNKNRNSKLSALQLSGARDN	300

Qy	301	AYTSAIECLLRFDVDRGLIIQYKNLSLTSTPQPKLSLVNITVNFISINAVSLVTTQD	360
Db	301	AYTSAIECLLRFDVDRGLIIQYKNLSLTSTPQPKLSLVNITVNFISINAVSLVTTQD	360

QY 361 IVDYQNSLSA 370
|||||
Db 361 IVDYQNSLSA 370

RESULT 2

US-10-688-058-64
; Sequence 64, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-64

Query Match 97.2%; Score 1837; DB 5; Length 370;
Best Local Similarity 97.0%; Pred. No. 7.9e-140;
Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPQDTISVSLDSRIQASRPNTYNPLLVYKTAKIKVKNDAANYKILNLTNNYKQIETL 60
|||||
Db 1 MPQDTISVSLDSRIQASRPNTYNPLLVYKTAKIKVKNDAASYKILNLTNNYKQIETL 60
|||||
QY 61 EKDNNGDQFCKEYKTLKTAMSNFFNSSEESLKSADLFYKDKPEELKYLKVRHTFV 120
|||||
Db 61 EKENGNGEDQFCKEYKTLKTAMSNFFNSSEESLKSADLFYKDKPEELKYLKVRHTFV 120
|||||
QY 121 VLINTEGNSDGLKIYKDDYDKFTPSIFFVFSKQEIKELFKDKGKTEKERNIAVYS 180
|||||
Db 121 VLINTEGASDGLKIYKDDYKFKPSTFFVFSKQEIKELFKDKGKTEKERNIAVYS 180
|||||
QY 181 NNKDNLHLKFISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKURTAKNFYSLNETG 240
|||||
Db 181 NNKDNLHLKFISQYLHQASIFHAVNPYGMPLAATPLVDDTVIGKURTAKNFYSLNETG 240
|||||
QY 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
|||||
Db 241 LDGVPAPKEGVDLAGGAIDQFTYHYIKNEAIIELIRIWNKNRQNSKLSALQLSGARDN 300
|||||
QY 301 AYTSAEICLLKRFVDRGLIIQYKNLSLTSPQPKLELSVNITNFNSINAVSLVITQD 360
|||||
Db 301 AYTSAEICLLKRFVDRGLIIQYKNLSLTSPQPKLELSVNITNFNSINAVSLVITQD 360
|||||
QY 361 IVDYQNSLSA 370
|||||
Db 361 IVDYQNSLSA 370
|||||

RESULT 3

US-10-688-058-62
; Sequence 62, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: JOHNSTON, STEPHEN A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058

; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 62
; LENGTH: 160
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-62

Query Match 41.7%; Score 788; DB 5; Length 160;
Best Local Similarity 98.1%; Pred. No. 1.3e-55;
Matches 157; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 211 LAATPLVDDTVIGKURTAKNFYSLNETGLDGVPAKFGVDLAGGAIDQFTYHYIKNE 270
|||||
Db 1 LAATPLVDDTVIGKURTAKNFYSLNETGLDGVPAKFGVDLAGGAIDQFTYHYIKNE 60
|||||
QY 271 AIIELIRIWNKNRQNSKLSALQLSGARDNAYTSAIECLLKRFVDRGLIIQYKNLSLTLS 330
|||||
Db 61 AIIELIRIWNKNRQNSKLSALQLSGARDNAYTSAIECLLKRFVDRGLIIQYKNLSLTLS 120
|||||
QY 331 PTPQPKLELSVNITNFNSINAVSLVITQDIDVYQNSLSA 370
|||||
Db 121 PTPQPKLELSVNITNFNSINAVSLVITQDIDVYQNSLSA 160
|||||

RESULT 4

US-10-688-058-18
; Sequence 18, Application US/10688058
; Publication No. US20050058661A1
; GENERAL INFORMATION:
; APPLICANT: SYKES, KATHRYN F.
; APPLICANT: HALE, KATHERINE S.
; APPLICANT: JOHNSTON, STEPHEN A.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR VACCINATION COMPRISING
; TITLE OF INVENTION: NUCLEIC ACID AND/OR POLYPEPTIDE SEQUENCES OF THE GENUS
; TITLE OF INVENTION: BORRELIA
; FILE REFERENCE: MCRO:003US
; CURRENT APPLICATION NUMBER: US/10/688,058
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-688-058-18

Query Match 17.4%; Score 329; DB 5; Length 65;
Best Local Similarity 100.0%; Pred. No. 4.3e-19;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 LAATPLVDDTVIGKURTAKNFYSLNETGLDGVPAKFGVDLAGGAIDQFTYHYIKNE 270
|||||
Db 1 LAATPLVDDTVIGKURTAKNFYSLNETGLDGVPAKFGVDLAGGAIDQFTYHYIKNE 60
|||||
QY 271 AIIEL 275
|||||
Db 61 AIIEL 65
|||||

RESULT 5

US-10-282-122A-70437
; Sequence 70437, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel


```
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: 09/386,962
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-690-184-4

Query Match          6.9%; Score 130.5; DB 4; Length 1742;
Best Local Similarity 22.1%; Pred. No. 0.44;
Matches 95; Conservative 59; Mismatches 169; Indels 107; Gaps 21;

QY 18 SRPNYNNPLLVYKTAKIKVVK-DAANYKILNLTNNYKEQIETLEKONGNQDQFGKEKT 76
Db 122 SKPKANEAVVNESTKPKTTEAPTNEESIAETPKTSTTQODSTEKNPSSLKDNLSNST 181
QY 77 LLKTAMSNFFNSSESLKSADLIYK-----DKPEELKKYLKVRHRTFVVLINTEGDSN 130
Db 182 TSKE-----KTDEHSTKQAMSTNKSNDLTNDSTQSEK-----TSSQANDSTNQ 229
QY 131 D-----DGLKIYKDDYD-----KFKTPSIFPVSTKEQE-----IKE 162
Db 230 SAPSKQLDSKPSKQVYKTKFNDEPTQDVEHTTTKLKTPSVSTDSVNDKQDYTRSAVAS 289
QY 163 LFKDKGNTKERNIAVYNNKNDLHLKFS-OYLHOASIFHAV-----NP-YGM--PLAA 213
Db 290 LGVDSNETE-----AITNAVRNDLDKAAREQINEALIAEALKKDFSNPDYGVDTPLAL 344
QY 214 TPLVDDTVIGKLTAKINFYSLNLTGL-----DGVPAFKEGVDL-----AGGAI----- 258
Db 345 NRSQSKNSPHKSPRMLMSLAEPNSGKNVNDVKITNPTLSLNKSNHANNVWPTS 404
QY 259 DEQTY--HYIKNEAI-----IELIRWNKNRQNSKLSALQLSGAR 298
Db 405 NEQFNLKANYELDDSIKEGDTFTIKYQYIRPGGLELPAI-----KTQLRSKDGSI VANGVY 461
QY 299 DNAYTSAIECLLKRFVDRGLIIQYKNL--SLTSPQLKLELSVNIYFNFSINAVSLVI 356
Db 462 DKT-TNTTYYTFTNYVD-----QYQNTGSPDLIATPKRETAIKONQYPMVEVTIANEVV 515
QY 357 TTQDIVDYQN 366
Db 516 KKDFIVDYN 525

RESULT 8
US-10-689-082-4
; Sequence 4, Application US/10689082
; Publication No. US20040142348A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy
; TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES FROM COAGULASE-NEGATIVE STAPHYLO
; FILE REFERENCE: P06335US04/BAS
; CURRENT APPLICATION NUMBER: US/10/689,082
; CURRENT FILING DATE: 2003-10-21
; PRIOR APPLICATION NUMBER: 09/386,962
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/098,443
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/117,119
; PRIOR FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1742
; TYPE: PRT
```

```
; ORGANISM: Staphylococcus epidermidis
US-10-689-082-4

Query Match          6.9%; Score 130.5; DB 4; Length 1742;
Best Local Similarity 22.1%; Pred. No. 0.44;
Matches 95; Conservative 59; Mismatches 169; Indels 107; Gaps 21;

QY 18 SRPNYNNPLLVYKTAKIKVVK-DAANYKILNLTNNYKEQIETLEKONGNQDQFGKEKT 76
Db 122 SKPKANEAVVNESTKPKTTEAPTNEESIAETPKTSTTQODSTEKNPSSLKDNLSNST 181
QY 77 LLKTAMSNFFNSSESLKSADLIYK-----DKPEELKKYLKVRHRTFVVLINTEGDSN 130
Db 182 TSKE-----KTDEHSTKQAMSTNKSNDLTNDSTQSEK-----TSSQANDSTNQ 229
QY 131 D-----DGLKIYKDDYD-----KFKTPSIFPVSTKEQE-----IKE 162
Db 230 SAPSKQLDSKPSKQVYKTKFNDEPTQDVEHTTTKLKTPSVSTDSVNDKQDYTRSAVAS 289
QY 163 LFKDKGNTKERNIAVYNNKNDLHLKFS-OYLHOASIFHAV-----NP-YGM--PLAA 213
Db 290 LGVDSNETE-----AITNAVRNDLDKAAREQINEALIAEALKKDFSNPDYGVDTPLAL 344
QY 214 TPLVDDTVIGKLTAKINFYSLNLTGL-----DGVPAFKEGVDL-----AGGAI----- 258
Db 345 NRSQSKNSPHKSPRMLMSLAEPNSGKNVNDVKITNPTLSLNKSNHANNVWPTS 404
QY 259 DEQTY--HYIKNEAI-----IELIRWNKNRQNSKLSALQLSGAR 298
Db 405 NEQFNLKANYELDDSIKEGDTFTIKYQYIRPGGLELPAI-----KTQLRSKDGSI VANGVY 461
QY 299 DNAYTSAIECLLKRFVDRGLIIQYKNL--SLTSPQLKLELSVNIYFNFSINAVSLVI 356
Db 462 DKT-TNTTYYTFTNYVD-----QYQNTGSPDLIATPKRETAIKONQYPMVEVTIANEVV 515
QY 357 TTQDIVDYQN 366
Db 516 KKDFIVDYN 525

RESULT 9
US-10-732-923-22524
; Sequence 22524, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22524
; LENGTH: 1877
; TYPE: PRT
; ORGANISM: Plasmodium yoelii
US-10-732-923-22524

Query Match          6.9%; Score 129.5; DB 5; Length 1877;
Best Local Similarity 21.4%; Pred. No. 0.59;
Matches 83; Conservative 51; Mismatches 122; Indels 131; Gaps 19;

QY 9 SLDSRQASRP-----NYYNPLLVTAKIKVKNKDAANYKILNLTNNYKEQIETLEK 62
Db 1238 TLKDSRI---RPIILTGDNAYNCLYV-----GNKIGLFNNNTYE-SFYSLVN 1281
QY 63 DNGNGQDQFGKEKTLTKTAMSNFNSSESLKS-----ADLFYKCKPEELKKYLKVRH 116
Db 1282 NSNNIDHSNKKKN-----TNFSNNNSFNLSDKFMPFENLFKYSSSKPKFNKNVNO 1335
QY 117 HTFVVLINTEGDSDDGLKIYKDDYDKFTPSIFPVSTKEQEIKELFKDKGNTKERNI 176
```



```

Db      1336  -----IYDNNNDNDD-----NDND-----DDNDNDNDNN 1362
Qy      177  AVYSNNK-----DNLHLKFIQYLHQASIFHAVNPYGMPLA-----AT 214
Db      1363  NSNSNNNNGDDDNNGVAKKQKRSKYNTINTINQYGSNILDIDASKIEBYLLESYKMN 1422
Qy      215  PLVDDT-----VIGKLRITAKINPYSLLNETGLDV-----PAPKEGYDLAGGAIDEQ 261
Db      1423  PVNDTNYDQNIIVGYGYLLNELIFYNIHNDKNINNTIVLYKDIYKE-IILTGEA---- 1477
Qy      262  FTHYIKNEAIIELIRIWNKNRNSKLSALQLSGARDNAVTSIAECLLKRFFVDRGLIIQ 321
Db      1478  --YNYIRD---HIFKIQNEFENIS-----NLHHTNNNT-----FKQNPFEQ 1517
Qy      322  YKNLSLTL-----SPTQKLUKLELSVNIT 344
Db      1518  YKNFLKVRIFSRULTPNKMEIIKDFT 1544

RESULT 10
US-10-732-923-22547
; Sequence 22547, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22547
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(1879)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-22547

```

RESULT 11

US-10-724-972A-7483

Sequence 7483, Application US/10724972A

Publication No. US20040147734A1

GENERAL INFORMATION:

APPLICANT: Doucette-Stamm, Lynn

APPLICANT: Bush, David

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: PATH03-16

CURRENT APPLICATION NUMBER: US/10/724,972A

CURRENT FILING DATE: 2003-12-01

PRIOR APPLICATION NUMBER: 09/450,969

PRIOR FILING DATE: 1999-11-29

PRIOR APPLICATION NUMBER: 09/134,001

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 7544

SEQ ID NO 7483

LENGTH: 575

TYPE: PRT

ORGANISM: S.epidermidis

US-10-724-972A-7483

Query Match 6.4% Score 120.5; DB 4; Length 575;

Best Local Similarity 20.6% Pred. No. 0.61;

Matches 84; Conservative 68; Mismatches 131; Indels 125; Gaps 22;

Qy 29 YKTAKIKVNKDA-ANYKILNLTVNNYEKQIETLEKNGN-GODQFGKEKTLTKTAMS--- 83

Db 69 YKAWKNKFDQVQKGAKSMNIAPIVQKRNVKIDKNGNIVRDNNGNPKTERKPVITGYR 128

Qy 84 --NFFNSBSLSKADLFIYKDKPEELKYLKVRHTFVVLINTEGNSDDGLKIYKODY 141

Db 129 AHNVDVAD--TKGKPLITAKD-----LIKTEFENSND---YKDL 164

Qy 142 DKK-----TBSI---FV-----FSTKEQEI-----KELFKDK 167

Db 165 NEFQYINEELTSPVEEKHFLDPTLSNGAKYSPKSDIEIVSDDLSYDMRFKTLIHEY 224

Qy 168 GNTKEKRNIAVSNKNLHLKFTISOYLHOASIFHAVNPYCMPLAATPL-----V 217

Db 225 AHSQHN--DIGKTOISEHSRSIKETEASSAVVANYGLDTSYSLGVLSCWGHNIS 282

Qy 218 DDTVIGKLRITAKINFYSLNLET--GLDGVPAFKEGVDLAGAIDEQFTYHYIKNEAIEL 275

Db 283 DDELKAHIK--NIHSFAKTTIEEINSLPEFSQYID---KLESEM-----NKDVYKD 329

Qy 276 IRTWKNKRON--SKLSALO-----LSGARDNATSAIECLLKRFVDREGLLIYQKNL 325

Db 330 LSTMTIDLNLKNGFDKITIIRKGNLVNDYGLNEISENSYEND-----DFKVNIDYKGF 380

Qy 326 SLTSLPTPOLKLEI-----SVNITYNFS-----INAVSLVITTD 360

Db 381 N-TNNSODQAKIELINKHDDSLARDYNTFTQTYNRVNINNTTIFVQDD 427

RESULT 12

US-10-282-122A-54216

Sequence 54216, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

```

; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54216
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-10-282-122A-54216

Query Match          6.4%; Score 120.5; DB 4; Length 598;
Best Local Similarity 22.5%; Pred. No. 0.65;
Matches 97; Conservative 69; Mismatches 133; Indels 133; Gaps 25;

QY 10 LLDRIQASRPNYPLLVYKTKAKVKNDAANYKILNLTVN---NYEKQIETL----- 60
DB 171 LLD--LKATFPN-LKADFFKAVDNVFKIINKISVSKNLNHNLENTHTTTTHSMQ 227
QY 61 ---EKDNGNQDQFGKEKTLTKTAMSNFPN--SSEESLSKADLFYKDKPEELKKYLKVH 115
DB 228 KTNKDSG-----SLLSQTLKNLDSILSSKES-----KHEKNDKVKG--KIE 267
QY 116 RHTFVVLINTEGNSDDGLKIYKDDYDKFTPSIFVFSSTKEQ-----EIKELPKD---K 167
DB 268 EDT-----TDAKNTLNKIK--NDEFKAKNLTEEL-NIKDKNQDNLNKEKDLNKPDK 318
QY 168 GNTKERNIAVYNNKND-NLHLKFTISQYLHQASIFHAVNPYGMPLAATPLVDDT--VIGK 224
DB 319 LNKNGEKNLNQENTQDQKNLKNNDQNLNLDKNLN-----KEIVKDTQKLKSN 367
QY 225 LRTAKINFVSLNETGLDGPAPFKEGVDLAGGAIDEQFTYHYIKNEAIELIRIWNK--- 281
DB 368 LQAKDFN---LNKEPKNN--NKENKDIKQNFQDKLNFENLNKTQVVQ-----NKNN 415
QY 282 ----NNRQNSK-----LSAQLSGARDNAYTSATEC 308
DB 416 ANFNNTNNKNTKFTQEQTKTHSENVVDKNSLDNLSVKNLKNVQTNQNNARNITPKETLQY 475
QY 309 L---LKRFRVDRGLIIOYK---NLSLTLSPTQLKLELSV-----NITNPSINAVSLVI 356
DB 476 FQSDLEKAVD-----QYKAPITKLSITLNNLGEVEVTLIQGNLHNFNSNANML 530
QY 357 TTQDIDVDYQNSL 368
|: : : : :|

Db 531 FIONQAEFKNSL 542

RESULT 13
US-10-282-122A-72168
; Sequence 72168, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72168
; LENGTH: 1178
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72168

Query Match          6.3%; Score 119; DB 4; Length 1178;
Best Local Similarity 22.6%; Pred. No. 2.2;
Matches 67; Conservative 50; Mismatches 118; Indels 62; Gaps 16;

QY 14 RIQASRPNY-----NPLLVTAKIKVKNDAANYKILNLT--VNNYKQIETL-- 60
DB 250 KVKONLAAYQERDLLETKNQTLKEKRHQLSRQMDQKQADLLEMTLRLISDYERQIERIHL 309
QY 61 -----EKDNGNQDQFGKEKTLTKTAMSN---FFNSSEESLSKADLFYKDKPEELK 109
DB 310 EVSQTEKTKQTSQKSEQLSEQKASLKTFLAQKEVDLEQLEEDLKQKQEI--KNVETELS 368
QY 110 KY-----LKVHRHTFVVLINTEGNSDDGLKIYKDDYDKFTPSIFVFSSTKEQEIKE 162
DB 369 RFATDPDHIIESLREDFVRLMQKEADTSNQ-LAVLKAEMDSRKQES-----ESKTAETKQ 422
QY 163 LFKD--KGNTEKERNIAVYNNKNDNLHLKFTISQYLHQASIFHAVN-PYG-MPLAATPLVD 218
DB 423 VQADLEKAKREQRESANFETAKTKVQ--ELLKDYQKTAQLVQNLLEAAYAEQOEAMFQLLD 481
```


GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 24, 2006, 19:59:25 ; Search time 5.92175 Seconds
(without alignments)
633.203 Million cell updates/sec

Title: US-10-688-058-20
Perfect score: 1889
Sequence: 1 MPQDTISVSLDSRIQSRP.....AVSLVITTDIVDYQNSLSA 370

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 70609 seqs, 10134256 residues

Total number of hits satisfying chosen parameters: 70609

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US12_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	132.5	7.0	US-10-793-626-1780	Sequence 1780, Ap
2	107	5.7	US-10-793-626-38	Sequence 38, Appl
3	103.5	5.5	US-10-467-657-2208	Sequence 2208, Ap
4	102.5	5.4	US-11-051-453-42	Sequence 42, Appl
5	102	5.4	US-11-052-554A-91	Sequence 91, Appl
6	100.5	5.3	US-11-051-453-41	Sequence 41, Appl
7	98	5.2	US-11-055-822-1066	Sequence 1066, Ap
8	96	5.1	US-10-512-109-27	Sequence 27, Appl
9	95.5	5.1	US-10-793-626-440	Sequence 440, App
10	94	5.0	US-11-069-642-20	Sequence 20, Appl
11	94	5.0	US-10-485-517-141	Sequence 141, App
12	94	5.0	US-11-013-759-11	Sequence 11, Appl
13	93.5	4.9	US-11-155-288-20	Sequence 20, Appl
14	93	4.9	US-11-055-822-1070	Sequence 1070, Ap
15	93	4.9	US-10-513-786-1	Sequence 1, Appl1
16	93	4.9	US-10-513-786-2	Sequence 156, App
17	92.5	4.9	US-10-873-528-156	Sequence 76, Appl
18	92.5	4.9	US-11-196-475-76	Sequence 1230, Ap
19	92.5	4.9	US-10-793-626-1230	Sequence 2594, Ap
20	92.5	4.9	US-10-793-626-2594	Sequence 786, App
21	92	4.9	US-10-793-626-786	Sequence 62, Appl
22	92	4.9	US-11-074-176-62	Sequence 66, Appl
23	91.5	4.8	US-11-196-475-66	Sequence 69, Appl
24	91	4.8	US-11-052-554A-69	Sequence 145, App
25	91	4.8	US-10-485-517-145	

ALIGNMENTS

RESULT 1

US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1

; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1780
; LENGTH: 1155

; TYPE: PRT
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match	7.0%	Score 132.5;	DB 6;	Length 1155;
Best Local Similarity	22.3%	Pred. No. 0.042;		
Matches	96;	Conservative 58;	Mismatches 169;	Indels 107;
Gaps	21;			
QY	18	SRPNYNPLLVYKTAIKVKNK-DAANYKILNLTNNYKQIETLEKONGGQDQFKEKKT 76		
Db	113	SKFKANEAVVTNNESTPKPTTEAPTVEESIAETPTSTTQDSTKNNPSLKDNLNLSST 172		
QY	77	LKLTAMNPFNSSESLSKADLFYK-----DKPEELKKYLVKVRHFWVLNTEGDSN 130		
Db	173	TSKES-----KTDEHSTKQAOQSTNKNLDTNDSPTQSEK-----TSSQANNDSTDNQ 220		
QY	131	D-----DGLKIYKDDYD-----KFKTPSIFVFSTKEQE-----IKE 162		
Db	221	SAPSKQLDSKPEQKVKYTKFNDEPTQDVEHTTKLTPSISTDSVNDKQDYTRSAVAS 280		
QY	163	LPFKDGNTEKERNIAVYNNKDNHLKFTS-OYLHOASIFHAV-----NP-YCM-PLAA 213		
Db	281	LGVDNSET-----AITNAVRNLDLKAASRQINEAIIAEALKDQSNPDYGVDTPLAL 335		
QY	214	TPLVDDTVIGKLTAKINFYSLNTEGL-----DGVPAFKEGVDL-----AGGAI----- 258		
Db	336	NTSQSKNSPHKASPRMNLMSLAEPNSGKNVNDKVKITNPITLSLKNNNHANNWIWPTS 395		
QY	259	DEQFTY--HYIKNEAI-----TELIRWNNRQNSKLSALQISGAR 298		

[illegible]

```

RESULT 2
US-10-793-626-38
; Sequence 38, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 38
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-38

```

Query Match	5.7%	Score	107;	DB	6;	Length	544;		
Best Local Similarity	22.4%	Pred. No.	1.1;						
Matches	71;	Conservative	51;	Mismatches	95;	Indels	100;	Gaps	17;

QY	18	SRPNYNNPLL--VYKTAIKIVKNDAAANYKILNLTIVNNYEQIETLEKDNNGOD-----Q	70
Db	2	SMENHIERLINHVKEITK-----EYAFSLGKSNIKAKVKLLKPKNYLRDITKEIQ	55
QY	71	FGKEHTLLKTAMGNFNNSSSELSKSDLFIYKDKPEEL-----KKYL-----	112
Db	56	KPQKQTGAPSWVKIDIVTEE-----EVLTFKDVKDELQTRRNYIDFGIALQYWNLSF	110
QY	113	----KVRHPTFVVLINTEG-----DNSDDGLKIY-----KDDYD-----KFKTPS	148
Db	111	LPBEINTNAPIKPVKTGDKTKLILSEQINNYLRKYTNHKKKPAYDFYENKEVIKFKTKG	170
QY	149	IPFVFSTKEQEIKELPKDKGKTEKERNIAVYNNKONLHLKFTSOYLHQASIEHAVNPG	208
Db	171	-----FILDEQKIYEL-HDEGYYKGLRKV-----DYLH-KEIDQLIESGYF-----LG	212
QY	209	MPLAAT-----PLVDDTVIGKLRITAKINFYSLLNETG-----LDGVPAFKGEVDLAG	255
Db	213	NMLSDTCGRYQGYFPHPD-----KEINFYNILRHASSTVALIEGLDYLGEDLTIVE	263
QY	256	GAIDQFTVHYIKNEAI	272
Db	264	KAINYYIENFYFDNEG	280

RESULT 3
US-10-467-657-2208
; Sequence 2208, Application US/10467657
; Publication NO. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega

```

; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103434.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 2208
; LENGTH: 867
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2208

```

Query Match	5.5%;	Score 103.5;	DB 6;	Length 867;
Best Local Similarity	18.5%;	Pred. No. 3.5;		
Matches	85;	Conservative 70;	Mismatches 185;	Indels 119; Gaps 20
Qy	9	SLLDSRIQASRPNNYPLLVYKTAK--IKVNDAAANYKILNLTVN-----NYEKQ	56	
Db	32	TVVKSRLTVEPQAGBPLVLDGSAKLLSVKINGAAADYVLEGETLTITADVPSEPTVEVE	91	
Qy	57	IETLEKDNNGQDQFGKEKTLTKLTAMSNFPNSSB-ESLKSADLFIYKDKPBELKKY--L	112	
Db	92	TELLPAEN-----KSLMGLVASGNGLFTQCEPEGEFRKITFYI--DRPDMSKFTTTI	141	
Qy	113	KVRHHTFVVLINT-----EGDNSDGLKIYKDDYDKFPTSPFFVFSTKEQBKELFKDKG	168	
Db	142	VADKKRYPVLLSNGNKIDGGESDGRHWKWE-DPPAKPSYLFALVAGDLAVTE--DRF	197	
Qy	169	NTEKERNIAV--YSNNKDNLHLKFIQYLHQASIFHAWN---PYGMLAATPLVDDTVIG	223	
Db	198	TTMSGENVKIEFTYTADPKVGFVAVESLKNAMKWDTRFGLCYDLDFIMVAVAGDFNMG	257	
Qy	224	KLRTAKINFS-----LLNETGLDGYV-----AFKEG	250	
Db	258	AMENKGLNIFNTKFLVADSRATATDTPFEGESVVGHYFHNWGNRVTCRDMFQLSLKEG	317	
Qy	251	V-----DLAGGAI-----DRQF-----TWHYIKN-----	269	
Db	318	LTVFRDQFESGDRAGVARRIENIRLLRQNFPEPAGTAHPVPSVEENNFTMTVY	377	
Qy	270	EALIELIRIWNKRNQNSKLSALQLSGARDNAVYTAIECLLKRPVD-RGLIIQYKNLSLT	328	
Db	378	EKGAEVVRMYHTLLGEEGFQKGMKLYQRHDGQAVTCDDFRAAMADANGINLQDFALWYS	437	
Qy	329	LSPTPOLKLE--LSVNITYNFSINAVSLVITTDQIDVYQ	365	
Db	438	QAGTPVLEAGRLKNV---FELTIKQTVPTPTDMADQK	473	

RESULT 4
US-11-051-453-42
Sequence 42, Application US/11051453
Publication NO. US20050287150A1
GENERAL INFORMATION:
APPLICANT: AMBROSINO, DONNA
APPLICANT: BABCOCK, GREGORY J.
APPLICANT: BROERING, THERESA
APPLICANT: GRAZIANO, ROBERT
APPLICANT: HERNANDEZ, HECTOR JAVIER
APPLICANT: LOWY, ISRAEL
APPLICANT: MANDELL, ROBERT
APPLICANT: MOLRINE, DEBORAH
APPLICANT: THOMAS, JR., WILLIAM D.
APPLICANT: ZHANG, HUI-FEN
TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
USES THEREOF
FILE REFERENCE: MJ1-001
CURRENT APPLICATION NUMBER: US/11/051,453
CURRENT FILING DATE: 2005-02-04
PRIOR APPLICATION NUMBER: 60/542,357

PRIOR FILING DATE: 2004-02-06
PRIOR APPLICATION NUMBER: 60/613,854
PRIOR FILING DATE: 2004-09-28
NUMBER OF SEQ ID NOS: 82
SOFTWARE: Patent in Ver. 3.3
SEQ ID NO 42
LENGTH: 2367
TYPE: PRT
ORGANISM: Clostridium difficile
US-11-051-453-42

Query Match 5.4%; Score 102.5; DB 7; Length 2367;
Best Local Similarity 20.6%; Pred. No. 15;
Matches 89; Conservative 65; Mismatches 153; Indels 125; Gaps 20;

QY 2 QDTSVSLDSRIQASR-PNYNPL-----LVYTKTIKV-----NKDAA 41
DB 603 PYDSI-----LFORNIEDSEVAIYNPTDSEIQEIDKYRIPDRISDRPKIKLTFIGHGKAEF 659
QY 42 NYKIL-NLTVNNYERQIET---LEKDNQGDQFGKEKTKLLKTAMSNFNFSESLKSAD 97
DB 660 NTDIFAGLDVLSLSEIETAIKLAKEDISPK---SIEINLLGCNMFSYVNVVEETYPGKL 716
QY 98 LFIYKDKPEELKXYLK-----VHRHTFVVLINTEG-----DNSDDGLKIYKDDYDKFKTP 147
DB 717 LLRVKDKVSELMPSQSDSIIVSANQYEVRIINSEGRRELLDHSGEWIN-----KEE 767
QY 148 SIFFVFSTKE-----QEIKEPLKDKNGTKERNIAVYNNKDNHLKFIQSYLHQAIF 201
DB 768 SIIRKDISKEYISFNPKNKIIVSKNLPSELSTLLQEIIRNNSNSDIE-----815
QY 202 HAVNPGMPLAATPLVDVDTVIGKLTAKINFYSLINTEGLDGVPAFKEGVDLAGAIDEQ 261
DB 816 -----LEEKVM--LAECEINVISNIETQVVE--ERIEAKSLTSDSI---853
QY 262 FTHYIKNE-AIIE-----LIRIMKNNRONSLSALQSGARDNAYTSAIECLLKRFV- 314
DB 854 ---NYIKNEFKLIESISEALCDLQKQNELEDSHFISPEDISETDEGFSI-----REIN 903
QY 315 -----DRGLIIOYKN-----LSLTGSPTPQLKLELSVNITYNFSINAVSLV 355
DB 904 KETGESIFVETEKTIFFSEYANHITBEISKIKGTIFDTVNGKLKVKVNLDTTDEVNTLNA 963
QY 356 ITQDIDVYQNS 367
DB 964 FFIQSLIEYNS 975

RESULT 5
US-11-052-554A-91
Sequence 91, Application US/11052554A
Publication No. US2005028866A1
GENERAL INFORMATION:
APPLICANT: Sachdeva, et al.
TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
FILE REFERENCE: 30853/40359A
CURRENT APPLICATION NUMBER: US/11/052,554A
CURRENT FILING DATE: 2005-02-07
PRIOR APPLICATION NUMBER: US 60/589,227
PRIOR FILING DATE: 2004-07-20
PRIOR APPLICATION NUMBER: IN 173/DEL/2004
PRIOR FILING DATE: 2004-02-06
NUMBER OF SEQ ID NOS: 763
SOFTWARE: Patent in version 3.3
SEQ ID NO 91
LENGTH: 2902
TYPE: PRT
ORGANISM: Helicobacter pylori J99
US-11-052-554A-91

Query Match 5.4%; Score 102; DB 7; Length 2902;
Best Local Similarity 19.6%; Pred. No. 22;

Matches 81; Conservative 46; Mismatches 126; Indels 160; Gaps 17;
QY 12 DSRQASRPNYNPLLVYTKAKIKVNDAAANYKILNLTNNYERQIETLEK-----62
DB 2278 NAOQFANNLTISNOAVLEKNASFVTN---NLNTQGAFFNNATQKIEVLQNLVIASNASL 2333
QY 63 -----DNGNQDQFGKEKTL-----LKTAMSNFFNSSEES-----92
DB 2334 STGIYGLEVGALNMLGAIHFNLENSQTPVNPILQVGGIINLNTTQTPFMVSVANGTY 2393
QY 93 --LKSADLFYKDKPEELKXYLKVHRHTFVVLINTEGNSDD--GLKIYKDDYDKFKTPS 148
DB 2394 TLLKSSYIIDYINPNLSQSLKY-----TLININGHIEEKNGVLTLYCQ-----2440
QY 149 IFFVFSTKEQIEKELPKDKG-----NTEKERNIAVYNNKDNHLKFIQSYLHQA 198
DB 2441 -----RVLLQDKGLLSVALPNSNNAQQNLSLS-----VLHNQIKMSYGNKV 2484
QY 199 SIFHAVNPGMPLAATPLVDVDTVIGKLTAKINFYSLINTEGLDGVPAFKEGVDLAGG-- 256
DB 2485 MDF-----TPPTLQDYIVG-----IQQSALNQ-----IEAVGGNN 2515
QY 257 AI-----DEQFTYHYIKNEAIIELI-----RIMNKNRQ-----286
DB 2516 AIKWLSTMMETKENPLFAPYILENHSLEITLGYTKDLQNTASLISNPNFRNATSLLEM 2575
QY 287 -----SKLSALQSGARDNAYTSAIECLLKRFVDRG---LIQYKNLS 326
DB 2576 ASYQTSRLTKLSDFRAREGESNFSERLLELKNKRFSDPNPSEVFKYSQLS 2628

RESULT 6
US-11-051-453-41
Sequence 41, Application US/11051453
Publication No. US20050287150A1
GENERAL INFORMATION:
APPLICANT: AMBROSINO, DONNA
APPLICANT: BABCOCK, GREGORY J.
APPLICANT: BROERING, THERESA
APPLICANT: GRAZIANO, ROBERT
APPLICANT: HERNANDEZ, HECTOR JAVIER
APPLICANT: LOWY, ISRAEL
APPLICANT: MANDELL, ROBERT
APPLICANT: MOLRINE, DEBORAH
APPLICANT: THOMAS, JR., WILLIAM D.
APPLICANT: ZHANG, HUI-PEN
TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
FILE OF INVENTION: USES THEREOF
FILE REFERENCE: MJ1-001
CURRENT APPLICATION NUMBER: US/11/051,453
CURRENT FILING DATE: 2005-02-04
PRIOR APPLICATION NUMBER: 60/542,357
PRIOR FILING DATE: 2004-02-06
PRIOR APPLICATION NUMBER: 60/613,854
PRIOR FILING DATE: 2004-09-28
NUMBER OF SEQ ID NOS: 82
SOFTWARE: Patent in Ver. 3.3
SEQ ID NO 41
LENGTH: 2710
TYPE: PRT
ORGANISM: Clostridium difficile
US-11-051-453-41

Query Match 5.3%; Score 100.5; DB 7; Length 2710;
Best Local Similarity 19.7%; Pred. No. 25;
Matches 94; Conservative 63; Mismatches 180; Indels 141; Gaps 19;

QY 4 DTISVSLDSRIQAS-----RPNYINPLLVYTKAKIKVNDAAANYKILNLT 50
DB 682 DTIKLIDISPKNVNLLGCNMFSYDFNVEETYPGKLILSIMDKITSTLPDVKNSITIGA 741
QY 51 NNYEKQIETLEKDNQGDQ-----FGKEKTLTKTAMSN---FFNSSESLK-----94

Db 742 NOYEVRI-----NSEGRKELLAHSGKWINKKEAIMSDLSSEKYEYIFFDSIDNKLKAKSKN 795
QY 95 -----SADL-----PIYKDKPEELKKYLKVRH- 116
Db 796 IPGLASISEDITKLLLDASVSPDTKFILNNLKNLNISSIGDYIYKEKLEPVKN--IHNS 853
QY 117 -----HTFVVLINTG-----NSDGLKIYKDDYDK-FKTPSIFFVSTKQE1-- 160
Db 854 IDDLIDEFNLENVSDLEYELKLANLDBKYLISPEDISKNNSTYSVRFINKSNGESVVV 913
QY 161 ---KELF-----KDKGNTKERNIAVYNNKDNLHLKFISQYLHQASIFHAVNPVM 209
Db 914 ETEKEIFYSKEHITKEISTIKNSITDVGNNLNDNIQLDHTSQ-----VNTLNA 963
QY 210 PLAATPLVD---DVTIGKLRTA-KINFVSLINETGLDGPAPAFKEGVDLAGGAIDEQFTY 264
Db 964 AFFIOQLIDYSSNKVDLNDLSTSVKQLYAQIFSTGLNTIYDSIQLVNLSNAV----- 1017
QY 265 HYIKNEAIELIRIWNKNNRQNSKLSALQSGARDNAVYTSATECLLKRFVDRGLIIQYKN 324
Db 1018 ---NDTINVLPITTEGPIVSTILDGINL-GAAIKELLDEHDPLKKLEAKVGVLA 1072
QY 325 LSLTSPFPOLKLELSVNITY-----NFSINAVSLVIT-----TQDIVDYQNSLS 369
Db 1073 MSLSTAATVASIVGIGAEVITIFLLPIAGISAGIPSLVNNELILHDKATSVVNFNHL 1130

RESULT 7

US-11-055-822-1066
; Sequence 1066, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 1066
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-1066

Query Match 5.2%; Score 98; DB 7; Length 701;
Best Local Similarity 20.6%; Pred. No. 6.6;
Matches 84; Conservative 41; Mismatches 95; Indels 188; Gaps 22;

QY 1 MPQDTISVS-----LLDSRIQASRPNNYNPLLVYKTAKIKVVKD-AANYKI LNL 48
Db 354 MYEDTVNASNPIEGRIHTSNLCSHLOVSTSEFNDDLT-----AEVGEDISCNLSLV 409
QY 49 TV-----NNYEKOIET-----LEKDN-----GNGO-----DOFGKEK 75
Db 410 AMADAPNFEKTIETIRGLTAVPEQTSIDSVPSIRKNEAAHAIGLQMNHLHGFGRE- 468
QY 76 TLLKTAMSNFNSSEESLSKADLF-----LYKDKPEELKKYLKVRHRTFV 121
Db 469 -----HMYGSEALDFTNAYFAAVLYOCLRASNOIATERGERPKNF----- 510
QY 122 LINTEGMNSDGLKIYKDDYDKFKTPSIFFVSTKQE1KELFKDKGNTKERNIAVYSN 181
Db 511 -----ENSKYATGEYFDDFDAND-----FAPKSKVKELFA----- 541
QY 182 NKDNHLKFIQYQLHQASIFHAVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNLTGL 241
Db 542 -KSNIH-----TPTVEDWAALK-----ADVMEHGLFNR-NL 570
QY 242 DGVPAFKEGVDLAGGAIDEQFTYHYIKNEA-----IIEILIRIWNKNNRQNSKLSALQSG 296
Db 571 QAVP-----PTGSIS-----YINNSTSIHPIASKIBI-----RKEGKIGRVYPA 611
QY 297 AR-----DNAYTSAIECLL-----KRFVDRGLIIQYKNLSLTL 329
Db 612 PHMDNDNLEYFEDAYEIGYEKIIDIYAVATKYVDQG-----LSLTL 652

RESULT 8

US-10-512-109-27
; Sequence 27, Application US/10512109
; Publication No. US20050255546A1
; GENERAL INFORMATION:
; APPLICANT: KIRIN BEER KABUSHIKI KAISHA
; TITLE OF INVENTION: POLYPEPTIDE HAVING AN ACTIVITY TO SUPPORT PROLIFERATION OR SURVIV
; TITLE OF INVENTION: OF HEMATOPOIETIC STEM CELL OR HEMATOPOIETIC PROGENITOR CELL, AN
; TITLE OF INVENTION: FOR THE SAME
; FILE REFERENCE: 905WOIOP1572
; CURRENT APPLICATION NUMBER: US/10/512,109
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: US 60/376,001
; PRIOR FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 27
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-512-109-27

Query Match 5.1%; Score 96; DB 6; Length 829;
Best Local Similarity 19.7%; Pred. No. 11;
Matches 70; Conservative 45; Mismatches 119; Indels 122; Gaps 15;
QY 12 DSRIOASRPN-----YYNPLLVYKTAKIKVVKDAN-----YKILNLTVNNYEKOIETLEK 62
Db 204 DSRVALSTCNGLHGMFEDDTFVYMIETLELTDDEKSTGRPHIIQKTLAQYQSKQMKNLST 263
QY 63 DNGNGQDOF-----GKEKTLTKTAMSNFNSSEESLSKADLFYKDKPEELKKYLK 113
Db 264 D---GSDQWPLLPELQWLRRRKRAVNSRGVF-----EEMKYLELMIVNDH-KTYKKHRS 314
QY 114 VHRH-----TFVVLINTEGNSDDGLKIYKDDYDKFKTPSIFFV 152
Db 315 SHAHTNNFAKSVNVLVDSIYKEQLNTRVVLVAVETWTEKDHIDI-----TINPVQMLHD 368
QY 153 FSTKEQETKELFKDKGNTKERNIAVYNNKDNLHL-KFISQYLHQASIFH----- 202
Db 369 FSKYRQRIKQ-----HADAVHLISIRVTFPHYKRSLSVPGGVCSRIR 409
QY 203 --AVNPYGMPLAATPLVDDTVIGKLTAKINFYSLNLTGLDGVPAFK----EGVDLAGG 256

Db 410 GGVNEXGLPMA-----VAQVLSQSLAQLGQWEPSSRKPKECIESWGG 455
Qy 257 AIDEQFTVHYIKNEAIIELIRWNNRQNSKLSALQSGARDNAYTSAIECLLKR 312
Db 456 CIMEETGV-----SHSRKFSKCSILEY---RDFLQGGGACLFNR 492

RESULT 9

US-10-793-626-440
; Sequence 440, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 440
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-440

Query Match 5.1%; Score 95.5; DB 6; Length 258;
Best Local Similarity 20.0%; Pred. No. 2.7; Mismatches 90; Indels 73; Gaps 11;
Matches 52; Conservative 45;
Qy 73 KEKTLTKTAMSNFFNSSEBSLSADLFYKDKPEELKYLKVHRTFVVLINTEGDSDD 132
Db 42 KEAVSLQTSQEGN-IHTKEVNGK---FIYASK-QDIEKAMQI-----KHSND 82
Qy 133 GLKIYKDYDKPTPSIFVFSTKEQEKELPKQKNGTEKERNIAVYNNKONLHLKFS 192
Db 83 DLK-YMDISEK-----VPMSEKEVNHILKGILENGKSTFIKAQDKYEVNLYL- 131
Qy 193 QYLHOASIFHAVNPYGMPLAATPLVDVTIGKLTAKINFYSLNETGLDGVPAPKEGVD 252
Db 132 -----ISHA-----LVETGNGQSLSKGKEGNHHYNNF-----GIGAFDEDAV 171
Qy 253 LAGGAIDEO-----FTVHYIKNEAIIELIRWNNRQNSKLSALQLSG 296
Db 172 KTGKSFQKQKVTTEKAIMGGAWFVRHYFRNNQLSLYQMRWNPQNG-----QHQQY 224
Qy 297 ARDNAYTSAIECLLRFVDR 316
Db 225 ASDIQANNIADMEKYDK 244

RESULT 10

US-11-069-642-20
; Sequence 20, Application US/11069642
; Publication No. US20050260626A1
; GENERAL INFORMATION:
; APPLICANT: LORENS, JAMES B.
; APPLICANT: PRAY, TODD R.
; APPLICANT: KINGELLA, TODD M.
; APPLICANT: BENNETT, MARK K.
; TITLE OF INVENTION: IN VIVO PRODUCTION OF CYCLIC PEPTIDES FOR
; FILE REFERENCE: RIGL-022CIP3
; CURRENT APPLICATION NUMBER: US/11/069,642
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: 10/232,758
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 10/422,536

; PRIOR FILING DATE: 2003-04-23
; PRIOR APPLICATION NUMBER: 09/800,770
; PRIOR FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/187,130
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 168
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-11-069-642-20

Query Match 5.0%; Score 94; DB 7; Length 496;
Best Local Similarity 23.3%; Pred. No. 8.1;
Matches 57; Conservative 34; Mismatches 76; Indels 78; Gaps 12;
Qy 6 ISVSLDSRIQASRPNYVN-----PLAVYKTAKIKVKNKDAANYKILNLTNNYKQIETL 60
Db 274 IIFSLPESYINALTAGYFDTGCFSLLYDKAKKH-----NLRMVLTSKRRLDVL 322
Qy 61 EKDNNGQDQFQKEKTLTKTAMSNFFNSSEBSLSADLFYKDKPEELKYLKVHRTFV 120
Db 323 EK-IGVILNSIGILNLTLHKSREVYSLISNKSLET-----FKK---IAKYLKIRKEAFI 373
Qy 121 VLINTEGDSDDGLKIYKDDY-----DKFKTPSIF--FVFSTKEQE-----IKELF 164
Db 374 -----NGYKTYKHEHERFECDDLFPVKEVFKLTKEKGRKEILKDSKIHIENWY 422
Qy 165 KQKQNT---EKERNIAVYNNKONLHLKFSQ-----YHQASI 200
Db 423 KEKTNIPREKLTVLRVYANNSE--HKEFLEKIVNGDISFVRVKVENIPYDGYVDLSI 480
Qy 201 PHAVN 205
Db 481 KHNQN 485

RESULT 11

US-10-485-517-141
; Sequence 141, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 141
; LENGTH: 1290
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-141

Query Match 5.0%; Score 94; DB 6; Length 1290;
Best Local Similarity 22.4%; Pred. No. 28;
Matches 66; Conservative 46; Mismatches 91; Indels 92; Gaps 16;
Qy 37 NKDAANYK-----ILNLT-VNVEKQIETLEKONGN---GDDQFQKEKTLTKTAMSNFF 86
Db 63 NONVANAKTTAKNALNLTSTINNAQKALKSQIEGATTVAGVQVSTTASELNTAMSNLQ 122
Qy 87 NS-SEESLSADLFYKDKPEELKYLKVHRTFVVLINTEGDSDDGLKIYKDDYDKFK 145

Db 123 NGINDEAATKAAL-----NGTONLEK-AKOHANTAI-----DGLS----- 156
 QY 146 TSPISFFVSTKEQIKELPKD-----KGNTKEKERN-----IAVSNKNDLH 187
 Db 157 -----HLTNAQKALQVLQOSTTVAEQNEQKANNVDAAMDKLQSIADNATTKQON 211
 QY 188 LKFIQYLHQASIFHAVNPG-MPLAATPLVDDTVI-----GKLTAKINFYSLNMETGLD 242
 Db 212 YTDASQNKDAYNNAVTTAQGIIDQTSPTLPTVINQAAGQVSTK-----NALN--GNE 265
 QY 243 GYPAKEGVDLAGGIDQFTHYIKNEAIIELIRIWNKNNRONSLSALQLSGA 297
 Db 266 NLEAAQQAQSLSGLD-----NLNNAQQTVD-QINGA 299

RESULT 12
 US-11-013-759-11
 ; Sequence 11, Application US/11013759
 ; Publication No. US20050249747A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Sasaki, Ken
 ; APPLICANT: Yang, Yan Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
 ; FILE REFERENCE: 1038-921MIS:jb
 ; CURRENT APPLICATION NUMBER: US/11/013,759
 ; CURRENT FILING DATE: 2004-12-16
 ; PRIOR APPLICATION NUMBER: US/09/361,619
 ; PRIOR FILING DATE: 1999-07-27
 ; NUMBER OF SEQ ID NOS: 32
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 2314
 ; TYPE: PRT
 ; ORGANISM: Moraxella catarrhalis
 US-11-013-759-11

Query Match 5.0%; Score 94; DB 7; Length 2314;
 Best Local Similarity 22.5%; Pred. No. 61;
 Matches 75; Conservative 47; Mismatches 138; Indels 74; Gaps 16;
 QY 31 TAKIKVNDAAKYKILNTVNNYKQIETLEKXNGODQFGKXTLLKTA-----MSNF 85
 Db 795 TAKVYDETNQSKV-TYDVNDEKTIK-LTGDNGK-TNKGIVKTYTTTNNWAKATNF 851
 QY 86 FNSSELSKADLFYIKDKPEELKYLK-VHR-----HTFVVLINTEGNSDDGLK 135
 Db 852 STTNDALVNA-----KDAENLNTLAKETHTTGTADTALQTFKV--KKDGTADTET 904
 QY 136 IYKDDYDKFTPSIFVSTKQEIKELPKDKGNTKEKERNIAVYNNKNDLHLKFIQYL 195
 Db 905 VGKDTQNGTKYN-----TLKLG-----ENGLVATNKDGTVTFGINTQSG 946
 QY 196 HQASIFHAVNPGMPLAATPLVDDTVIGK--LRTAKINFYSLNTEGLDGVPAKE--G 250
 Db 947 LKAGSTTLNKDGLSIKPNASNEQIQVGADGVKFAKVGKN--SSTGIDGTSRITTKDQIG 1004
 QY 251 VDLAGGAIDQFTHYIKNEAIIELIRIWNKN-NRQNSKLSALQSGARDNAYTSAIECL 309
 Db 1005 FTGANGSLDIT-KPHLTCKLKVGEVETINTGINAGKKITNIQ----- 1047
 QY 310 LKRFVDRGLIIQKNSLSLTSTPQLKLSVNI 343
 Db 1048 -----SGDITQNSDAVTGGRVYDLKTELESKI 1075

RESULT 13
 US-11-155-288-20
 ; Sequence 20, Application US/11155288
 ; Publication No. US2006008468A1
 ; GENERAL INFORMATION:

; APPLICANT: Chiang, Chih-Sheng
 ; APPLICANT: Simard, John J.L.
 ; TITLE OF INVENTION: COMBINATIONS OF TUMOR-ASSOCIATED
 ; TITLE OF INVENTION: ANTIGENS IN DIAGNOSTICS FOR VARIOUS TYPES OF CANCERS
 ; FILE REFERENCE: MANK.050A
 ; CURRENT APPLICATION NUMBER: US/11/155,288
 ; CURRENT FILING DATE: 2005-06-17
 ; PRIOR APPLICATION NUMBER: 60/580,969
 ; PRIOR FILING DATE: 2004-06-17
 ; NUMBER OF SEQ ID NOS: 40
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 20
 ; LENGTH: 976
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-11-155-288-20
 Query Match 4.9%; Score 93.5; DB 7; Length 976;
 Best Local Similarity 22.9%; Pred. No. 21;
 Matches 50; Conservative 33; Mismatches 108; Indels 27; Gaps 6;
 QY 14 RIQASRPNNYNPPLVYKTAKIKVNDAAANYKILN---LTVNNYKQIETLEKXNGODQ 70
 Db 605 RKQVENKNKYTELQEQENKALK-KKGTABSQLANVYKVNKLELELESASAKQKFGTDT 663
 QY 71 FGKKTLLKTAMSNFFNSSELSKADLF-----YKDKPEELKYLKVRHRTFV 121
 Db 664 YKQIEDKISEENLLEVEKAKVIADAVKQKIDKRCQHKIAEMVALMEKHKHQYDK 723
 QY 122 LINTEGNSDDGLKIYKDDYDKFTPSIFFVST-----KQEIKELPKDKGNTK 172
 Db 724 II--EERDSGLYKSKQEQSSLRASLELSNLKAEILLSVKKQLEIEREKEKLEKREA 781
 QY 173 ERNIAVYNNKNDLHLKFIQYLHQASIFHAVNPGMP 210
 Db 782 KENTATLKEKDKKTKTQTFL---LETPEIYWKLDKSKAVP 816
 RESULT 14
 US-11-055-822-1070
 ; Sequence 1070, Application US/11055822
 ; Publication No. US20050260707A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Markus
 ; APPLICANT: Kroger, Burkhard
 ; APPLICANT: Schroder, Hartwig
 ; APPLICANT: Zeider, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
 ; FILE REFERENCE: BGI-121CPN
 ; CURRENT APPLICATION NUMBER: US/11/055,822
 ; CURRENT FILING DATE: 2005-02-11
 ; PRIOR APPLICATION NUMBER: 09/606,740
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: 60/141,031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: 60/142,101
 ; PRIOR FILING DATE: 1999-07-02
 ; PRIOR APPLICATION NUMBER: 60/148,613
 ; PRIOR FILING DATE: 1999-08-12
 ; PRIOR APPLICATION NUMBER: 60/187,970
 ; PRIOR FILING DATE: 2000-03-09
 ; PRIOR APPLICATION NUMBER: DE 19930476.9
 ; PRIOR FILING DATE: 1999-07-01
 ; PRIOR APPLICATION NUMBER: DE 19931415.2
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931418.7
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931419.5
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: DE 19931420.9
 ; PRIOR FILING DATE: 1999-07-08

· This Page Blank (uspto)

Matches 22; Conservative 11; Mismatches 26; Indels 16; Gaps 2;
QY 6 LVDD-----TVIGKLR TAKINFYSLN ETGLDGVPA FKGV DLAGG-----AID 49
Db 38 IVDDKSADAQKLA EVVSSERYSG LKINLV LSETK LKGAGARNRGIDLATGDYVCF LDADD 97
QY 50 EQFTYHYIKNEAII E 64
Db 98 EWHKDKLOQNLSLIE 112

Search completed: January 24, 2006, 19:59:15
Job time : 4.46592 secs

This Page Blank (uspto)